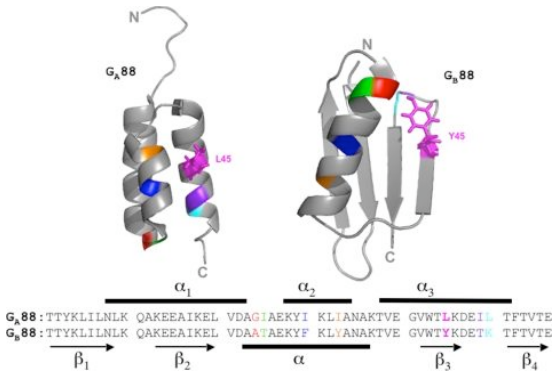
SAPIENZA
UNIVERSITÀ DI ROMA

DOTTORATO DI RICERCA IN BIOCHIMICA
CICLO XXV (A.A. 2009-2012)

Nearly identical, yet very different: comparing the folding of proteins with increasing degree of sequence identities but different structure and function



Dottorando
Rajanish Giri

Docente guide
Prof. Carlo Travaglini Allocatelli

Coordinatore
Prof. Paolo Sarti

Prof. Maurizio Brunori

Dicembre 2012

I heartily dedicate this thesis to my parents

Acknowledgements

This thesis owes its existence to the help, support and inspiration from many people across the world. It was a privilege and I am obliged and lucky to be one of the few foreign Ph.Ds of Sapienza University of Rome, Italy and work with a fantastic and glorious group of Prof. Maurizio Brunori, a great scientist, a critical thinker, and above all the greatest personality I ever met. I learnt a culture of incredible perseverance from my great mentor. Apart from the scientific point of view, he has been a guardian so that I enjoyed and lived these years fearlessly.

I am grateful to Prof. Carlo Travaglini Allocatelli, a real gentle man, for being my inspiring tutor and giving me full support of all kind. I am deeply indebted to Carlo Sir for taking up all my issues starting from Italian Visa.

I am very thankful to Prof. Stefano Gianni, a joyous and intelligent person who guided my work. Stefano Sir see the chevron plots!!

Above all getting three mentors together has developed many outstanding qualities to make me mature in all the areas, which I was destined for.

I would like to thank the coordinator of the Ph.D program in Biochemistry. In spite of my poor knowledge of Italian language Prof. Paolo Sarti and Prof. Francesco Malatesta has been very kind during all the three years.

I am thankful to Prof. Philip N. Bryan and Prof. Valerie Daggett for their precious contribution to our project.

My visit was always in excitement with the fabulous friend circle of Dr Angela Morrone (a very caring person in my group together with her meticulous style of work, she deserves a special appreciation), Eva, Luca, Linda, Flavio, Fulvio, Valentina, Giorgo, Serena and Carlo as being good friends who made a pleasant ambience.

I am deeply indebted to my parents who provided me spiritual support and forever standing byfor my all the good and bad initiatives. I heartly dedicate my thesis to my parents for a pure blind trust and support that was always extraordinary.
At the last but not the least, I am grateful to Dr Neha Garg who has been a lovely women in my life to bear me throughout this Ph.D and lifelong.

TABLE OF CONTENTS

INTRODUCTION

INTRODUCTION

1.1 The folding problem:

Protein folding is the process whereby the amino acid sequence (sometimes refers to as primary structure of proteins) is converted to a folded form, which exerts its biological function. The recent decades in the field of protein folding has been focused towards only a handful of model proteins and protein families comprising different sequences but same fold and function. Genome projects are pouring a substantial amount of new proteins in the databases on a daily basis and these new proteins are making a challenge to the current paradigm of studies done so far. One example is the knot proteins. Diversity of proteins found naturally and artificially evolved are itself making the protein-folding field an evergreen one.

Seminal discovery of Anfinsen on folding of ribonuclease demonstrated that primary structure of a protein encodes required information for a nascent polypeptide to fold into its native, physiologically active, three dimensional conformation (Anfinsen, et al. 1961). Generally small domain proteins are able to unfold upon addition of urea or other denaturants and can refold back upon removing the denaturants. The spontaneity of the reaction proves the concept that the native state is the most stable one corresponding to lowest free energy (Pace 1990).

After these discoveries the major question tackled was the time scale of the folding of a protein. Levinthal, taking the example of a 100 amino acid hypothetical peptide, showed that random search of all conformational states would take a very long time (10^{83} years), while practically the proteins fold in nanoseconds to seconds range (Levinthal, 1968; Levinthal 1969). "The Levinthal Paradox" put a type of foundation stone for the 'Protein Folding problem'.

From an experimental point of view, many proteins have been considered a model system to study the protein folding e.g. Barnase, RNase H, Chymotrypsin Inhibitor, PDZ domains, cytochrome C proteins (Fersht, 2008; Gianni et al, 2003; Brunori et. al. 2012; Travaglini et. al, 2009). Later on this study was focused mainly on protein families e.g. S6 family proteins (Lindberg and Oliveberg, 2007) and Ankyrin protein family (Ferreiro et. al, 2005)

While earlier studies focused on different proteins sharing the same fold (e.g. protein families), the recent availability of proteins with astonishing

hingh sequence identity (up to 98%), but different fold, provided for the first time the opportunity to study the folding process with a new and complimentary perspective. Ga88 and Gb88 are the most striking examples of heteromorphic proteins allowing to study the folding mechanism in a complimentary perspective (Morrone et. al, 2011). These two artificially evolved proteins (Alexander et. al., 2008) share a very high sequence identity i.e. 88%, while retaining different fold and function. Moreover, it has been shown that Ga88 and Gb88 have different folding mechanisms also starting from different denatured states. This work has prompted us to carry out a detailed folding study at molecular level, which I shall present and discuss in this thesis.

Another part of study was focused on Intrinsically Disordered Proteins (IDPs). These proteins are unfolded in physiological conditions and most of them are able to fold only upon binding their targets, IDPs represent therefore a very interesting system to study. In particular, the last part of the thesis will be focused on the study of the coupled folding and binding mechanisms of Intrinsically disordered protein transactivation domain of cMyb with its interacting partner KIX.

1.2 In-vitro folding studies:

1.2.1 The Role of equilibrium studies:

Equilibrium studies provide information about the folding process in terms of stability and co-operativity. The stability of folded proteins can be expressed as a linear function of the denaturant concentration (see inset panel of Fig. 1.1) (Tanford 1968). Consequently, in order to obtain the thermodynamic stability of a protein, expressed by the change in free energy between the native (N) and the denatured (D) state ($\Delta G_{D\text{-}N}$), the linear extrapolation method is routinely used. In these studies, the protein is unfolded with a chemical denaturant and a spectroscopic signal, generally fluorescence or circular dichroism, is recorded as a function of denaturant concentration. Following this approach, curves with sigmoidal shape are typically obtained (Figure 1.1). By applying simplified assumptions on the effect of a denaturant (Myers, et al. 1995), protein stability is estimated in the absence of denaturant (Pace 1986).

Moreover, this analysis allows to obtain a parameter, called $m_{D\text{-}N}$ value, which describes the co-operativity of the process. The $m_{D\text{-}N}$ value is the slope of the variation of the $\Delta G_{D\text{-}N}$ with denaturant concentration. On a purely empirical basis this value is correlated with the number of residues

and the surface area exposed to solvent during unfolding (Tanford 1968). $m_{D\text{-}N}$ value is a constant and is expressed in kcal mol^{-1} M^{-1}.

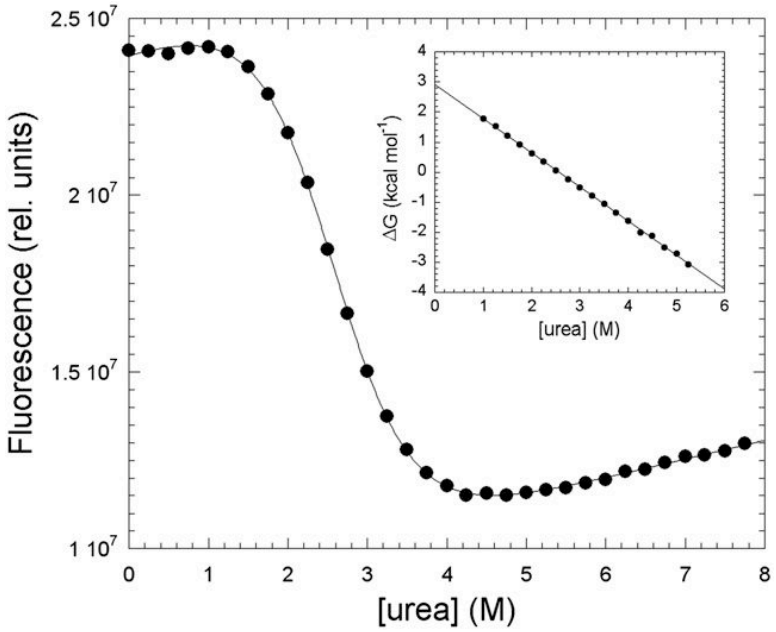


***Figure 1.1**. Equilibrium unfolding of the second PDZ domain from PTP-BL (PDZ2) monitored by fluorescence (Gianni, et al. 2005). Inset Panel: Linear free energy extrapolation. A quantitative analysis of the observed spectroscopic signal as a function of denaturant allows to estimate the stability of the protein in the absence of denaturant and to calculate the $m_{D\text{-}N}$ value, which describes the co-operativity of the process.*

1.2.2 The role of kinetics:

Folding studies of proteins are done by considering the unfolded state, as a primary reactant and the folded one as a final product, like a chemical reaction. The kinetic approach provides detailed information about the folding pathway of a protein and the reaction dynamics. Kinetic studies are based on rapid perturbations of the equilibrium conditions that change the energetics of the system. The perturbation induces a relaxation process to a new equilibrium estimating the relaxation rate of the reaction.

Small domain fast-folding proteins often exhibit two-state folding behavior (Jackson and Fersht 1991). Due to the fast folding of these small proteins, stopped-flow mixing devices with dead times in the millisecond range are

usually required, and the folding reaction is monitored by following the change in circular dichroism or Trp fluorescence signal. Folding and unfolding rate constants at a variety of denaturant concentrations are determined by fitting the folding and unfolding curves to exponential equations. In the case of two-state folding, only a single exponential folding phase is observed (Figure 1.2), whereas multi-state systems may lead complex kinetics.

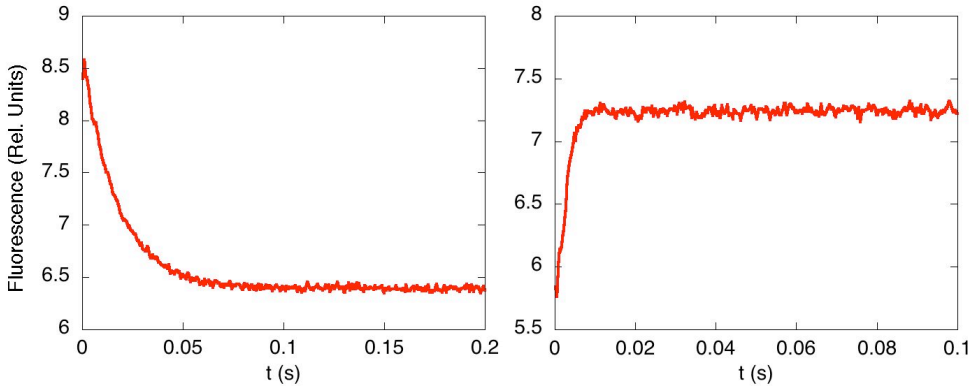


Figure 1.2. Stopped-flow time courses of the engineered protein G_A88. Unfolding (left panels) and refolding (right panels) reactions were monitored by the change in tryptophan fluorescence at pH 7.2 and 10 °C. The excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. The unfolding transition was initiated by a jump from urea 0 to 7.27 M; the refolding transition was initiated by a jump from urea 7 to 1.10 M. Both the curves are described by a single-exponential function.

If a protein folds following a two-state model (scheme 1.1), only the denatured and the native states are significantly populated (Buchner and Kiefhaber, 2005).

$$D \underset{k_U}{\overset{k_F}{\rightleftharpoons}} N \qquad \text{(Scheme 1.1)}$$

When a perturbation is imposed on the system, the observed rate constant k_{obs} is represented by the following equation:

$$k_{obs} = k_F + k_U \qquad \text{(Eq. 1.1)}$$

where k_U and k_F represent the unfolding and refolding rate constants respectively. In a two-state reaction, the logarithm of the observed (un)folding rate constants is linearly dependent on denaturant concentration (Jackson and Fersht 1991). Figure 1.3 shows the folding/unfolding rate constants expected for a two-state system. Due to the similarity of the goat horn 'chevre' with these V shaped plots it is called Chevron plot. From the analysis of a chevron plot (a) k_F^{H2O} and k_U^{H2O} are calculated which represent the extrapolated folding and unfolding rate constants in absence of denaturant and (b). The parameters m_F and m_U reflect the denaturant concentration dependence and correlate with the change in accessible surface area between the two ground states and the transition state in between.

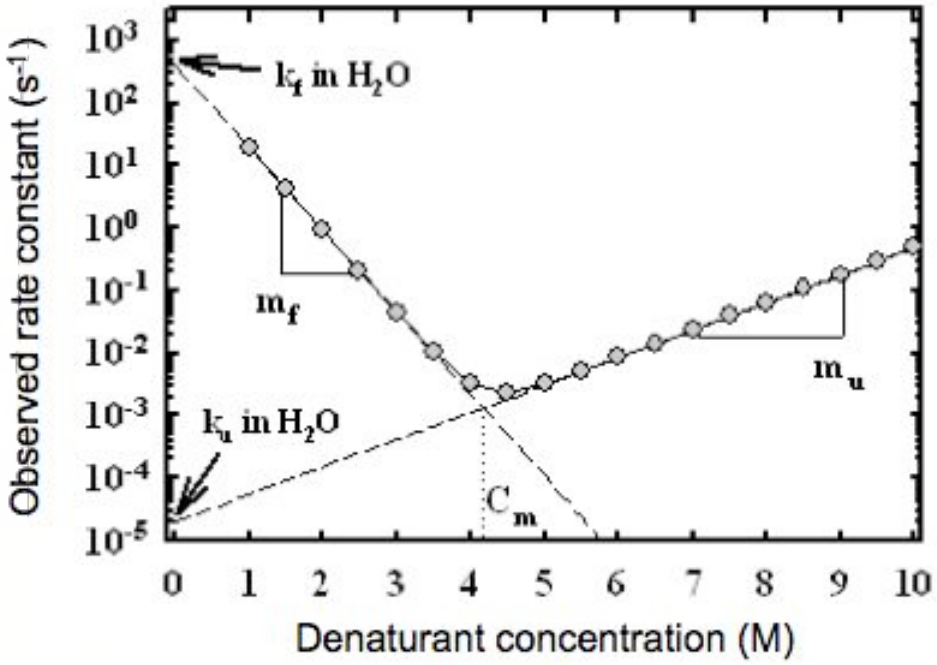


Figure 1.3. *Semilogarithmic plot of the observed rate constants as a function of denaturant concentration (chevron plot); simulated data set. Dashed lines represent the best fit to a two-state model. As described in the text, for a two-*

state system, the observed rate constant is the sum of the folding and unfolding microscopic rate constants.

1.3 Dissecting the folding pathway: Transition State Theory

Protein folding can be described as a unimolecular chemical reaction, in which the "reactant" (an unfolded protein) is converted to a "product" (the folded state). Unimolecular chemical reactions are typically governed by a single rate-limiting step, when the system passes through a high free-energy barrier called "transition state (TS)". The presence of two well-defined, thermodynamic macro states in a folding reaction implies the existence of an energy barrier in between. So two-state folding transitions (Figure 1.4) are described by the transition state theory (TST).

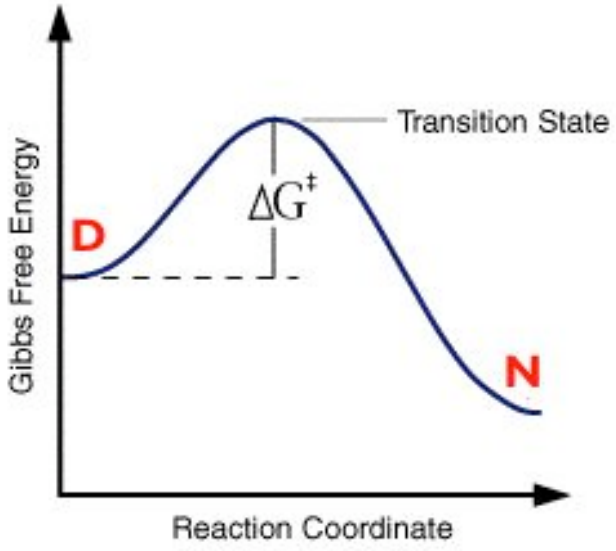


***Figure 1.4**. Transition State Theory (TST). Schematic illustration of free energy profile for a folding reaction. The transition state is located between the denatured state D and the native (folded) state N of the protein. The folding rate constant is proportional to the exponential of the (negative) activation free energy, the free energy difference between the transition state and the denatured state D.*

Transition state theory was first proposed by Eyring in 1935 to explain chemical reaction rates (Eyring 1935). Equation 1.2 is used to describe this theory:

$$k = \left(\frac{\kappa K_b T}{h}\right) e^{(-E_a/RT)} \qquad \text{(Eq. 1.2)}$$

where k is the rate constant of the reaction, κ is called transmission coefficient, K_b is Boltzmann's constant, h is Plank's constant, E_a is the activation energy for the reaction, R is the universal gas constant and T is the absolute temperature respectively. The term ($\kappa K_b T/h$) is a pre-exponential factor interpreted as the vibrational frequency, which refers to the maximal theoretical reaction rate ($\sim 10^{13}$ sec^{-1}) for simple molecular reactions. According to Eyring's equation the observed folding rate constant, k_f, is proportional to the free energy of activation (i.e., the free energy difference per mole between the denatured and the transition states, $\Delta G^{\ddagger}$).

One standard extension of the TST assumes that the reaction progress can be described by a reaction coordinate. In essence, an ideal reaction coordinate would be any degree of freedom that connects reactant**(**s**)** and product**(**s**)** along the lowest free energy continuous path on the free energy surface of the reaction; the highest free energy point along this path is the transition state (Figure 1.4). Moreover, while in ordinary chemical reactions, such as bond breaking in the gas phase, the TS corresponds to a well defined molecular structure. In protein folding the TS is described as ensemble of high free energy conformations; it is therefore referred to as the transition state ensemble (J. N. Onuchic 1996; J.N. Onuchic 2004).

1.4 Protein Folding Intermediates

Understanding the role and structure of partially folded intermediates is of fundamental mechanistic importance for protein folding studies. Earlier work suggested that the folding of small single domain proteins generally conforms to an all-or-none behavior (Jackson 1991), involving simultaneous formation of secondary and tertiary structure (Itzhaki, et al. 1995). Following this view, folding occurs in a two-state fashion, via condensation around a marginally stable nucleus, and discrete intermediates tend to be avoided

(Fersht 1995). When the inherent stability of folding nuclei is increased, however, even very simple protein systems appear to fold in a more complex fashion, with population, for example, of a partially folded intermediate, which may either transiently accumulate leading to multi-phasic kinetics, or be a high energy species en-route to the native state (Gianni, et al. 2003). In both cases, the system is represented by a three-state model (scheme 1.2).

$$D \underset{k_{ID}}{\overset{k_{DI}}{\rightleftarrows}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftarrows}} N \qquad \text{(Scheme 1.2)}$$

The presence of such local minima in the landscape is very difficult to address experimentally (Bryngelson, et al. 1995) and intermediates may sometimes escape detection.

1.4.1 Identification of Folding Intermediates: The roll-over effect

The analysis of chevron plots is a common and powerful tool for detecting protein folding intermediates. A deviation from its classical V-shaped appearance may indicate that the system under analysis is kinetically more complex than a simple two-state reaction. This deviation, known as roll-over effect, in either folding or unfolding branches of chevron plots, may result from different scenarios, including accumulation of intermediates (Figure 1.5 a) (Matouschek, et al. 1990; Parker, et al. 1995; Ferguson, et al. 1999) or changes in the rate-limiting step between two discrete barriers, due to a high-energy never accumulating intermediate (Figure 1.5 b) (Oliveberg 1998; I.E. Sanchez 2003; Gianni, et al. 2009).

It is sometimes possible to detect the accumulation of low-energy intermediates by observing multiphasic kinetics and/or by analyzing fluorescence amplitudes (Khorasanizadeh, et al. 1996; Ferguson, et al. 1999; Capaldi, et al. 2001; Jemth, et al. 2004). On the other hand, the mechanisms involving a change in the rate-limiting step (i.e., involving a high-energy intermediate) are more difficult to be detected. They often result kinetically equivalent to another event which could be responsible for a roll-over effect: a structural change in one single transition state, involving Hammond effects on smooth energy barriers (Figure 1.4 c) (Otzen, et al. 1999; Ternstrom, et al.

1999).

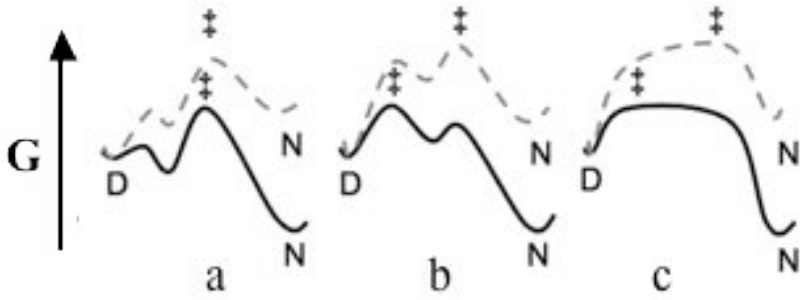


Figure 1.5. *Schematic free energy diagrams predicted for (a) three-state folding involving a low-energy accumulated intermediate; (b) three-state folding involving a high-energy never accumulating intermediate; (c) two-state folding involving a broad energy barrier. Solid black lines and dashed gray lines represent schematic energy diagrams for the two different scenarios expected at low and high denaturant concentrations, respectively. While the mechanisms involving a change in the rate-limiting step (b) or a smooth barrier (c) are generally indistinguishable, accumulation of a folding intermediate (a) results in double exponential time courses and may be addressed using ultrafast mixing/relaxation techniques (Shastry, et al. 1998; Capaldi, et al. 2001; Jemth, et al. 2004).*

A classical roll-over effect is present in the chevron plot reported in Figure 1.6. In this example, the observed rate constants follow a two-state behavior in the refolding and unfolding arms until 5.5 M guanidine concentration, whilst at higher denaturant concentrations, a deviation from linearity is detected.

The roll-over effect can be more or less pronounced depending on the difference between the *m* values of each state. In particular if the intermediate state is native like, i.e. highly compact, the roll-over can be described by a kink in the refolding branch. However in many cases, the folding intermediate can display unfolded-like properties; in such a case, the

deviation from linearity is much more difficult to detect and deviation from two-state folding may be detected by comparison between the thermodynamic parameters obtained by equilibrium and kinetic experiments (Matouschek, et al. 1989; Fersht 1999; Gianni, et al. 2003).

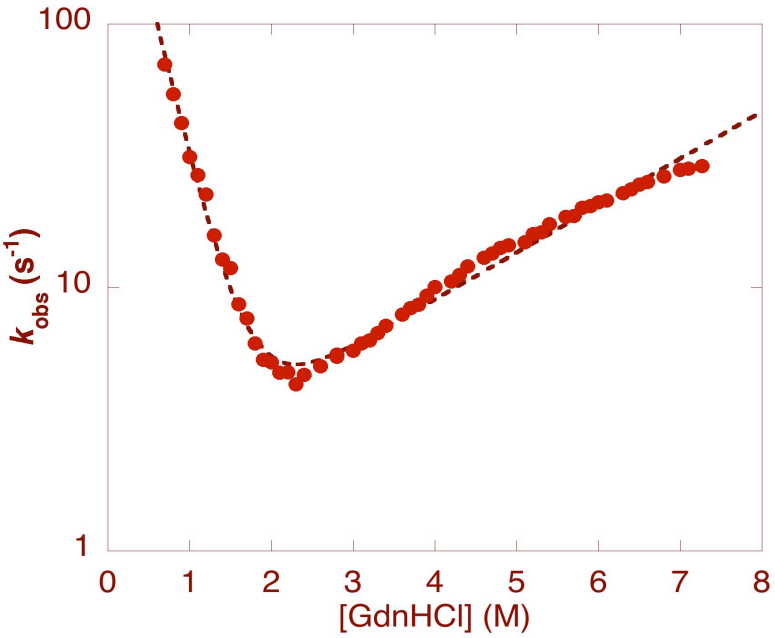


***Figure 1.6**. Chevron plot of the B1 IgG-binding domain of streptococcal protein G (generally called GB1) measured at pH 9.0 and 25 °. It is clearly evident that the data do not fit well to a two-state model (dashed line), since a deviation from linearity in the unfolding arm is present (roll-over effect).*

1.4.2 Alternative explanations for multi-state kinetics

Roll-over effects may result from events other than the presence of a high or low-energy intermediate state; thus analysis of chevron plots must consider all possibilities. In particular, a deviation from linearity in a chevron can result either from changes in the position of the transition state along the reaction coordinate or from transient aggregation of the denatured protein.

Changes in transition state ensembles consist in movements of the transition state position along the reaction coordinate as the denaturant concentration is increased. As shown in Figure 1.5 c, this instance is represented by a two-state model involving a broad energy barrier, assuming

the protein to display plastic folding pathways characterized by a malleable TS (Ternstrom, et al. 1999; Cellmer, et al. 2007; Lindberg and Oliveberg 2007;). In other cases, the TS is surprisingly robust and maintains its structural features when the system is perturbed, for example by altering solvent conditions or by mutagenesis (Jackson and Fersht 1991; Itzhaki, et al. 1995). In spite of the remarkable conceptual differences invoked by mechanisms implying either a malleable or robust TS, it is extremely difficult, if not impossible, to unequivocally discriminate between these two models (Scott, et al. 2004). This is not surprising, because these two models correspond to extreme manifestations of a more complex scenario, whereby folding is characterized by a rough energy landscape (Hyeon and Thirumalai 2003) as detailed in "*The Energy Landscape Theory*" section .

Another mechanism underlying complexity is due to the presence of transient aggregates i.e. association of two or more non-native protein molecules. This event driven by hydrophobic interactions, results in the formation of amorphous structures that lack long-range order. Because aggregation is sensitive to protein concentration, monitoring the kinetics as a function of concentration is a mandatory control to exclude aggregation artifacts (Figure 1.7).

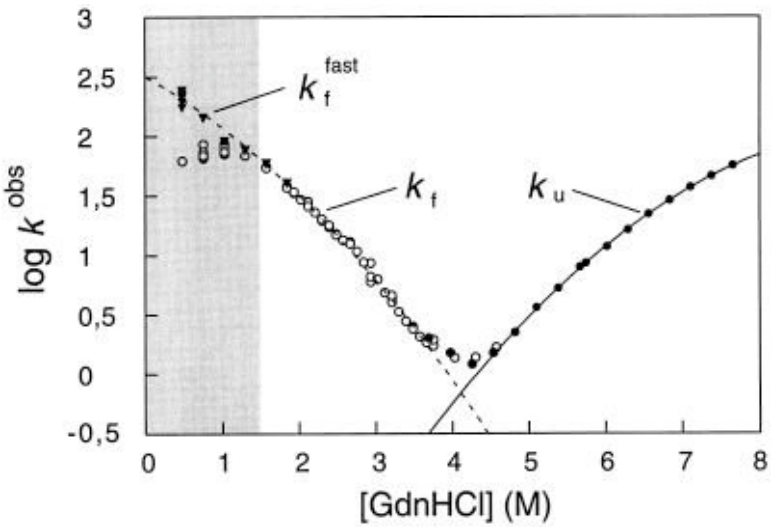


Figure 1.7. *An example of complexity due to aggregation and broad activation barrier. GdnHCl dependence of the rate constants for folding and unfolding of the spliceosomal protein U1A (Silow and Oliveberg 1997). The*

deviation from two-state folding observed at low [GdnHCl] (○), found also for other proteins, may be mistaken as accumulation of an intermediate. With U1A, this deviation was showed to be caused by transient aggregation of the denatured protein under refolding conditions (conditions where aggregation occurs are marked grey). At low protein concentrations the denatured protein remains monomeric during the refolding process while at higher protein concentrations the denatured protein aggregates in the dead-time of the stopped-flow instrument, giving rise to a retardation of the refolding rate. A part from transient aggregation commented above, it is evident the presence of a pronounced specular curvature in both the arms of U1A chevron plot. Silow and Oliveberg demonstrated that specular curvature is due to a broad activation energy barrier separating the native and the denatured states.

1.5 Experimental strategies for Folding studies

In analogy with classical organic chemistry, the best strategy to unravel the mechanism of the folding reactions would be to isolate all the intermediate states and characterize the transition states in between. Indeed, many advances in protein folding studies have been achieved by isolating intermediates and studying their structure. However isolation and characterization of folding intermediates is often impossible because they are short-lived and at very low concentration (due to cooperativity). Indeed, for many single domain proteins, only the fully native and the fully denatured states may be populated at equilibrium. In these cases, a key role in protein folding studies has been played by the description and the characterization of folding transition states. A great contribution is due to A. Fersht and co-workers (Fersht, et al. 1992) who introduced in the late 1980s the so-called Φ value analysis, contributing to the description of protein folding mechanism at nearly atomic resolution.

1.5.1 Φ–Value analysis

The protein engineering approach, termed Φ value analysis, was developed in the laboratory of Alan Fersht with the purpose of unveiling the structure of the protein folding transition states (Fersht, et al. 1992). Following this method, the degree of structure formation of individual

residues in the transition state is actually inferred from analyzing the effect of single-site mutations on folding rates and stability. The Φ value analysis has been extensively used to investigate the transition state of many proteins (Matouschek, et al. 1989; Gianni, et al. 2007a; Zarrine-Afsar, et al. 2010; Banachewicz, et al. 2011; Giri, et al 2012).
The Φ value is calculated as the ratio of the energetic perturbation induced on the transition state versus that induced in the native folded state, introducing a non-disruptive mutation, intended to cause a small perturbation:

$$\Phi = \frac{\Delta\Delta G^{*}_{TS-D}}{\Delta\Delta G_{N-D}} \qquad \text{(Eq. 1.3)}$$

where $\Delta\Delta G_{N-D}$ is the change induced by mutation in the free energy of folding and $\Delta\Delta G^{*}_{TS-D}$ that induced in the activation energy of folding.
Φ values normally range from 0 to 1. A Φ-value near unity indicates that the TS is energetically perturbed upon mutation as much as the native state. This effect indicates that in the TS the mutated residue is engaged in fully native contacts (i.e. it has all its native interactions established) (Figure 1.8 right panel). On the other hand, a Φ-value near zero is taken as evidence that the TS is not energetically perturbed by the mutation, while the native state is (i.e. the mutated residue is as unstructured in the TS as in the denatured ensemble) (figure 1.8 left panel).

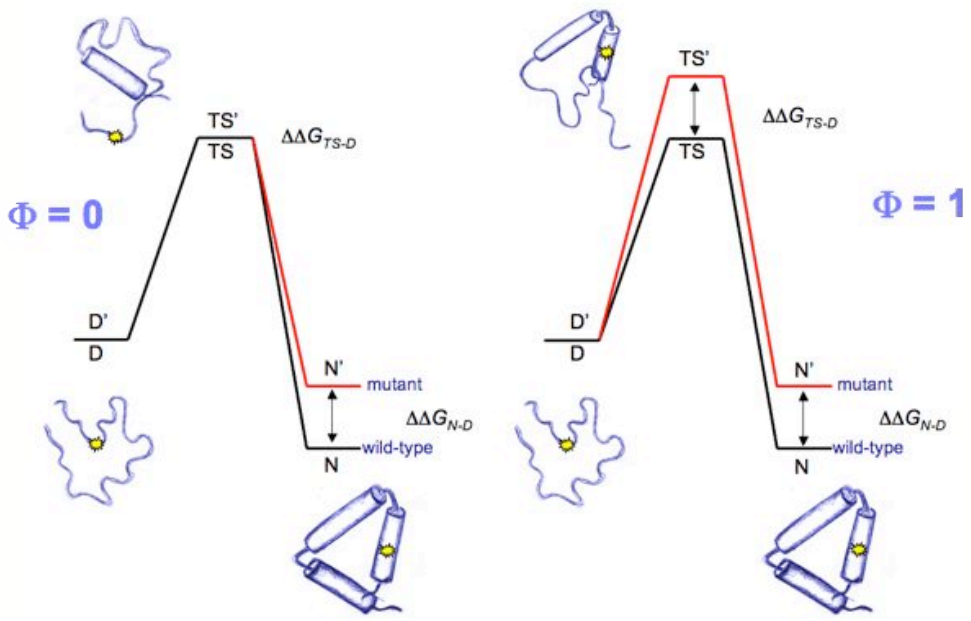


Figure 1.8. Φ-*value analysis. The depicted energy profiles represent two different scenarios in which a hypothetical probed residue (shown in yellow on the structures) gives rise to different effects in the folding activation energy and in the stability of the protein, depending on the structure of the transition state. Left panel: the site of mutation is in an unstructured region of the transition state (Φ=0). Right panel: the probed residue is highly native-like in the transition state (Φ=1).*

Experimentally determined Φ values are generally in between 0 and 1. The traditional interpretation of fractional Φ values is, however, not straightforward as they might indicate the existence of multiple folding pathways or a unique transition state ensemble with genuinely weakened interactions (Oliveberg and Fersht 1996).

Also the interpretation of the so-called non-classical Φ values (Φ> 1 and Φ < 0), that are seldom observed, is not straightforward. While, in some cases, these values are due to an incorrect mutation, in some others, they appear genuine. Negative Φ values can be observed when the native state is destabilized while the transition state is stabilized and vice-versa. Φ values higher than 1 may be detected when the transition state is affected by the

mutation more than that observed in native state (Gianni 2012). Because unusual Φ values are often indicative of non-native contact formation (Ozkan, et al. 2001), they can be used to detect local misfolding in transient intermediates either *en-route* to the productive folding pathway, as in the case of the Im7 protein (Capaldi, et al. 2002), or acting as off-pathway kinetic traps, like in the case of the circularly permuted PDZ domain of D1p protein (Gianni, et al. 2010).

1.5.2 Comparative studies on Protein Folding: Homologous Proteins

Current knowledge on the protein folding reaction has been achieved by extensively characterizing the folding mechanisms of simple globular proteins (Jackson 1998). However, given the diversity of protein folds and especially amino acid sequences, it is extremely difficult to draw general rules by studying folding pathways of individual proteins. In fact, when considering the folding of different proteins at least three key variables may jeopardize a comparison: i) sequence composition, ii) native and iii) denatured states structure variability.

A powerful strategy to elucidate some of the relationships between sequence information and folding mechanism is to study proteins that differ in sequence but share the same overall fold (Chiti, et al. 1999b; Clarke, et al. 1999a; Martínez and Serrano 1999; Riddle, et al. 1999; Friel, et al. 2003; Travaglini-Allocatelli, et al. 2003; Travaglini-Allocatelli, et al. 2005; Chi, et al. 2007; Calosci, et al. 2008). This strategy assumes that general correlations between amino acid sequences and folding pathways may be extrapolated by comparing folding processes for different members of a given protein family. The final goal is to identify the limited number of sequence determinants to achieve the common fold. Thus, the protein folding problem is often investigated using proteins with a low degree of sequence identity, yet adopting essentially the same fold.

Over the years, it has been found that the mechanism of folding is, generally, conserved in protein families (Chiti, et al. 1999a; Gianni, et al. 2001a; Zarrine-Afsar, et al. 2005) suggesting that the native topology is one of the main factors controlling protein folding (Baker 2000).

An interesting example of comparative study is represented by the cytochrome c (cyt c) family, which has been used extensively in different laboratories, as a model system for folding studies (Elove, et al. 1994; Bai, et

al. 1995; Colon, et al. 1996; Gianni, et al. 2001a; Brunori, et al. 2003; Travaglini-Allocatelli, et al. 2003). The c-type cytochrome are a large family of globular proteins with a characteristic α-helical fold and a covalently bound heme group. In 2004, using a comparative approach, it was shown that very different members of the cyt c protein family share a common folding mechanism (Travaglini-Allocatelli, et al. 2004). By comparing the folding kinetics of cytochrome c_{551} from the mesophilic bacterium *Pseudomonas aeruginosa* (Pa-cyt c_{551}) and cytochrome c_{552} from the thermophilic bacterium *Hydrogenobacter thermophilus*, it was demonstrated that the folding transition states of these two proteins share some similarities, in spite of large differences in thermodynamic stabilities. Analysis of extensive kinetic data available on many eukaryotic cyt c has indicated that similar species are populated along the folding pathway, enabling to propose a consensus mechanism.

The PDZ domains represent another example of comparative folding study. They constitute a large family of protein-interaction modules that mediate protein-protein recognition by binding to short amino acid sequences (Fanning and Anderson 1999). Different classes of PDZ domains recognize specific C-terminal sequences (PDZ motifs) on a variety of protein substrates. Analyzing and comparing the kinetic folding mechanisms of five related but distinct PDZ domains, it has been found that, despite their low sequence identity and apparent folding complexity, the folding reactions for PDZ domains can be explained by a model with an intermediate and two transition states that are rather conserved with regard to their positions along the folding reaction coordinate (Chi, et al. 2007). In particular, through a combination of Φ-value analysis and molecular dynamics simulations, it was found that the late transition states are much more structurally similar than the early transition states (Calosci, et al. 2008). Surprisingly, in a further study on a topological PDZ mutant (cpPDZ2), such as a circular permutant, where the native N- and C-termini were joined and the sequence cleaved in a different position (Ivarsson, et al. 2008), it was shown that, although circular permutation introduces a significant destabilisation of the native state, the folding kinetics of cpPDZ2 reveal a remarkable stabilisation of the folding intermediate, which accumulates transiently during folding.

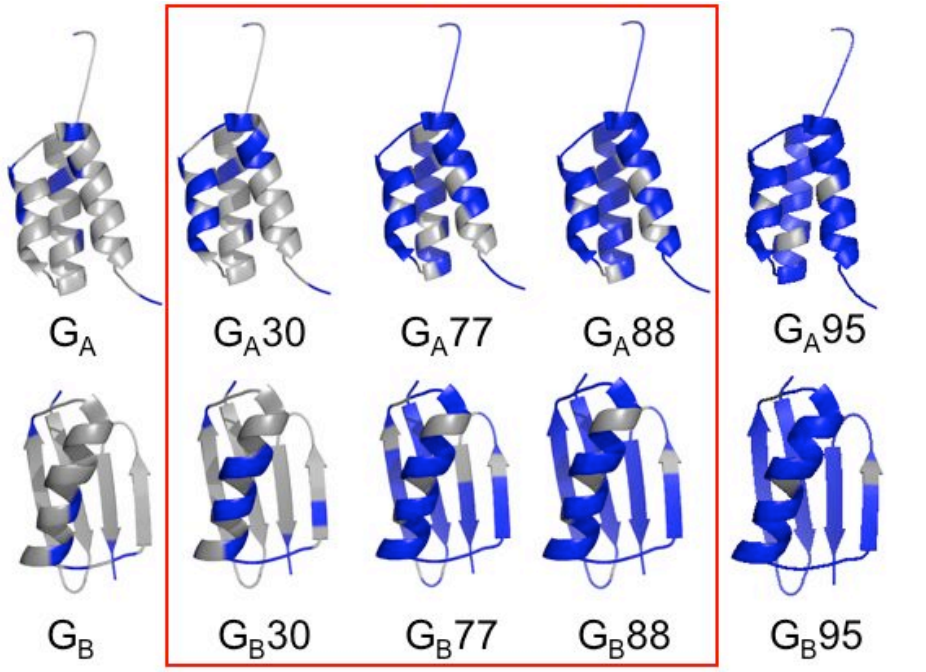
1.5.3 A complementary approach in Protein Folding studies: Proteins with high sequence identity but different fold

The analysis of heteromorphic proteins, that is proteins with high

sequence identity but different structure, can be considered as a novel and alternative approach complementary to the folding studies on protein families.

While the analysis of protein families allows to control one of the main factors governing protein folding, i.e. native-state topology, the study of heteromorphic proteins allows to control another relevant parameter: the amino acid sequence of a protein.

It is known that, proteins with a significant similarity in their amino acid sequence are expected to have the same fold. In fact, analysis of the protein data bank (PDB) reveals that a sequence similarity of 40% nearly always leads to a conserved three-dimensional structure and function (Wilson, et al. 2000). This observation provoked Rose and Creamer in 1994 to issue the ‘‘Paracelsus Challenge’’, whereby the protein folding community was charged with the task of designing two proteins that were at least 50% identical but possessed different folds (Rose and Creamer 1994). Amazingly, this goal was fully achieved in only 3 years, when Dalal and co-workers designed a sequence that, in spite of being 50% identical to a mostly β-sheet protein, folded into a four-helix bundle (Dalal, et al. 1997). Since then, several other scientists have achieved similarly impressive feats of design (Davidson 2008). In 2008, ambitious work by Bryan and co-workers led to the design of a pair of proteins with an extraordinarily high degree of sequence identity but different structure and function (Alexander, et al. 2007; He, et al. 2008). In particular, the sequences of two domains from streptococcal protein G were subjected to an iterative design of heteromorphic proteins leading the authors to produce three pairs of protein G variants with an increasing level of sequence identity (30%, 77% and 88% respectively) (Alexander, et al. 2007; He, et al. 2008) (figure 1.9). Two proteins, sharing 88% sequence identity (49 out of 56 amino acids), are called G_A88, which is mostly α-helical (the 3 helix bundle protein A fold), and G_B88, displaying the α+β protein G fold (Figure 1.9); yet they display divergent structures and functions that were similar to the respective wild-type proteins. The study of this remarkable protein engineering achievement offers some unique opportunities for a complementary analysis on protein folding mechanisms and allows to pose two key questions: (1) At which stage of its folding pathway does a protein commit to a given topology? and (2) Which residues are crucial in directing folding to a given structure?.




```
GA    MEAVDANSLA QAKKAAIKEL KQYGIGDYTI KLINNAKTVE GVESLKNEIL KALPTE
GB1   MTYKLILNGK TLKGETTTEA VDAATAEKYF KQYANDNGVD GEWTYDDATK TFTVTE

GA30  MEAVDANSLA QAKEAAIKEL KQYGIGEKYI KLINNAKTVE GVWSLKNEIL KALPTE
GB30  MTYKLILNGK TLKGETTTEA VDAATAEKYF KLYANDKTVE GEWTYDDATK TFTVTE

GA77  TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL KATVTE
GB77  TTYKLILNGK QLKEEAITEA VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE

GA88  TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL TFTVTE
GB88  TTYKLILNLK QAKEEAIKEL VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE

GA95  TTYKLILNLK QAKEEAIKEL VDAGTAEKYI KLIANAKTVE GVWTLKDEIK TFTVTE
GB95  TTYKLILNLK QAKEEAIKEA VDAGTAEKYF KLIANAKTVE GVWTYKDEIK TFTVTE
```

Figure 1.9. Structures and sequence alignments of the different G_A and G_B variants. All engineered G_A and G_B proteins display structure and function similar to their respective natural wild-type domains GA and GB1. For each protein, amino acid identities are shown in blue and nonidentities in gray. A complete folding characterization of the variants shown inside the red rectangle was performed.

1.6. Theoretical studies

A crucial approach in protein folding studies is represented by the comparison between experimental and theoretical results. Theoretical models and computer simulations have greatly advanced our understanding of protein folding and have helped the interpretation of experimental data. There is a synergy between theory and experiment, the former providing testable models and the latter aiming at validating hypothesis (Szilagyi 2007).

One of the principal tools in the theoretical approach to the structure and function of biological molecules is the method of molecular dynamics (MD) simulations (Karplus and McCammon 2002). Simulations can help in identifying or predicting transition and intermediate states along the folding pathway, they provide estimates of the rate of folding and in some cases, predict the final, folded structure.

Small proteins typically fold in the several microseconds to seconds timescale; detailed atomistic simulations, however, are currently limited to the nanosecond to microsecond regime. Therefore, simulation of folding requires either simplified models or special sampling methods, both of which introduce some approximations. To simulate unfolding, the simplest method of increasing sampling is to increase the temperature of the simulation to 500 K or more where the native structure of the protein is melting within a few nanoseconds. This approach has been applied to several small proteins, including bovine pancreatic trypsin inhibitor (Kazmirski and Daggett 1998b), lysozyme (Kazmirski and Daggett 1998a), myoglobin (Tirado-Rives and Jorgensen 1993), barnase (Wong, et al. 2000), ubiquitin (Alonso and Daggett 1998), the SH3 domain (Day and Daggett 2003), etc. Features of the unfolding process, such as the transition-state ensemble or the unfolded ensemble, have in some cases shown remarkable agreement with experimental results (Tsai, et al. 1999).

1.7 Folding mechanisms

In 1968 Levinthal postulated the existence of folding pathways. The huge diversity of proteins found in nature made it difficult to give a widely accepted unified scheme on folding mechanism. In 1973, the nucleation model (Wetlaufer 1973) tried to resolve Levinthal's paradox by assuming that the rate-limiting step in the folding process is a nucleation event, presumably the formation of smaller structural units, and once nucleation

occurs the nuclei grow fast and the folding process rapidly completes. The ‘‘diffusion–collision model (DC)” (Karplus and Weaver 1976) implies fluctuating microdomains (portions of secondary structure or hydrophobic clusters) that move diffusively and repeatedly collide with each other. Productive collisions lead to coalescence into intermediates, which may involve microdomains that are not necessarily contiguous along the protein sequence. Folding proceeds as a series of coalescence events that might either follow a unique order (sequential folding pathways) or explore different routes (parallel folding). Also the so-called ‘‘jigsaw puzzle model’’ (Harrison and Durbin 1985) denied the necessity of a unique, directed folding pathway and stated that each protein molecule can follow a different route to the native structure, just like there are multiple ways to solve a jigsaw puzzle.
The introduction of the ‘‘hydrophobic collapse model’’ (Dill 1985), based on the intuition of C. Tanford, led to the view that the hydrophobic effect is the main driving force of folding, and the process starts with a rapid collapse of the chain, expulsion of the water and formation of the secondary structure.

The ‘‘framework model’’ (Baldwin 1989) stated that the folding process is hierarchical, starting with the formation of the secondary structure elements, and the docking and the organization of the preformed substructures is the rate-limiting step.

The model named ‘‘nucleation–condensation (NC)’’ represents an attempt to unify the features of both the framework and the hydrophobic collapse mechanisms (Fersht 1995; Fersht 1997). In this model, secondary structure and hydrophobic interactions form nearly simultaneously and synergistically (Daggett and Fersht 2003), leading to the formation of a weak structured local nucleus. The nucleus is composed of a set of adjacent residues, stabilized by long-range interactions that are formed as the rest of the protein collapses around the nucleus: formation of the nucleus (nucleation) (Wetlaufer 1973) is coupled with a more extended formation of structure (condensation).

1.7.1 Nucleation-Condensation and Diffusion-Collision: Extreme manifestations of a common mechanism for Protein Folding

Among the models for protein folding, Nucleation-Condensation (NC) and Diffusion-Collision (DC) mechanisms are, currently, the two models generally used to describe the folding of small single-domain proteins. A critical test to distinguish between them is represented by the

analysis of the magnitude and distribution of Φ values for a given protein.
The transition state of a protein which folds according to the DC model, displays heterogeneous structure localization, with regions having Φ values close to 1 and others displaying Φ values close to 0, distributed in contiguous blocks, indicative of preformed secondary structure elements (or independent microdomains) (Gianni, et al. 2003).

On the other hand, in the case of the NC model, the nucleus may be identified by the few residues displaying higher Φ values; the native-like structure in the TS should gradually decrease with a smooth gradient of decreasing Φ values from the nucleus (Itzhaki, et al. 1995).

In 2007, a work on a PDZ domain (second PDZ repeat from Protein Tyrosine Phosphatase-Bas Like, PDZ2) (Gianni, et al. 2007a), performed in our laboratory, suggested that NC and DC models may represent extreme manifestations of an underlying common mechanism and that proteins may appear to fold by either NC or DC depending on the inherent stability of their secondary structure elements. As reported above, it was already shown that the folding pathway of PDZ2 proceeds through two consecutive TS barriers (Gianni, et al. 2005). The interactions formed in the two distinct TSs were mapped by Φ-value analysis. Surprisingly, the first TS was characterized by many mutants displaying Φ = 0 and only some mutants having fractional Φ values, suggestive of a NC mechanism. On the other hand, the second TS displayed characteristics of the DC mechanism with several Φ-values close to 1, and the rest displaying fractional Φ. The folding of PDZ2 has been suggested as a paradigmatic example in which the two extreme models are manifested in one single protein. According to this unifying folding mechanism, the folding of small globular proteins is suggested to involve three major events: (1) formation of a weak nucleus that determines the native-like topology of the structure, (2) a global collapse of the entire polypeptide chain, and finally (3) consolidation of the remaining partially structured regions to achieve the native state conformation (Gianni, et al. 2007a).

1.7.2 The Energy Landscape Theory

Since the process whereby a protein acquires its native three-dimensional shape involves formation of many non-covalent weak bonds, a classical one-trajectory view of protein folding is likely to be an over-simplification of the underlying mechanism. An original viewpoint emerged when the concept of energy landscape for a protein (Frauenfelder et al. 1991)

was extended to folding. Energy landscapes are mathematical devices that help to understand the microscopic behavior of a molecular system (Bryngelson, et al. 1995). Although energy landscapes are, by definition, high-dimensional surfaces, they are often pictured as a surface in three-dimensions. In these pictures, the vertical axis represents the free energy and the horizontal axes represent the conformational degrees of freedom of the polypeptide chain. The statistical energy landscape perspective describes folding as taking place on a rugged free-energy landscape in which free-energy barriers separate ensembles of states displaying different levels of structural heterogeneity, such as the folded (F) and the unfolded (U) states. This view likens the energy landscape of a protein to a funnel, with the native structure at its global minimum (Figure 1.10).

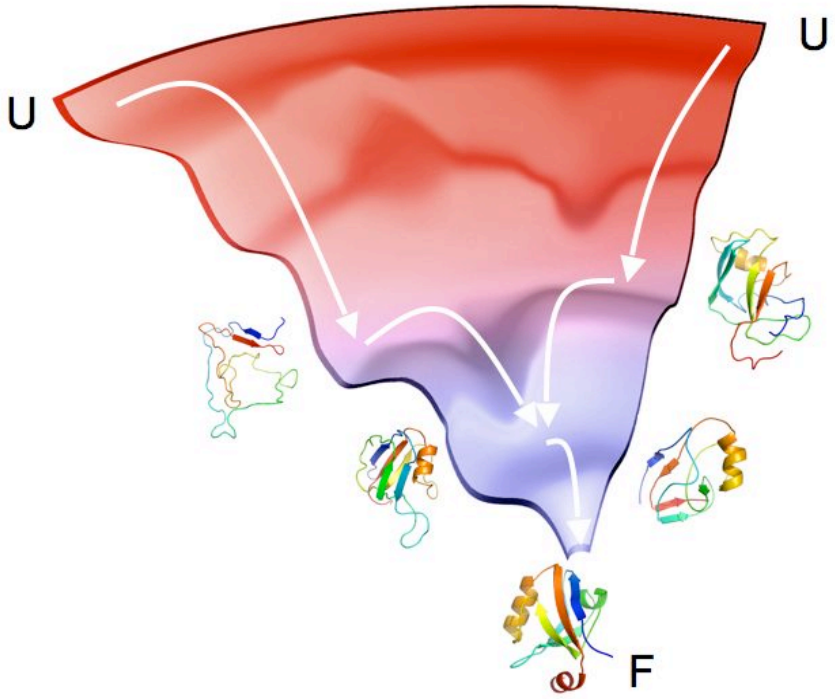


Figure 1.10. *Schematic representation of a folding funnel: the free energy drives the polypeptide towards the folded state (F), the most stable conformation, and the conformational entropy dramatically decreases as the native state is approached. Implicit in this view is that folding may occur via alternative parallel pathways. This concept is graphically depicted using the structures of the early and late folding transition states for different PDZ domains, as obtained in ref. (Calosci, et al. 2008).*

According to the landscape theory, the plasticity of folding pathways assumes that proteins can be rerouted through the energy landscape by mutational (Wright, et al. 2003), topological (Lindberg and Oliveberg 2007) or solvent perturbations (Gianni, et al. 2007b; Otzen and Fersht 1998). Thus, the polypeptide chain may fold by different pathways, potentially adopting multiple partially folded ensembles *en route* to the native state (J.N. Onuchic 2004). Alternative folding pathways, involving different nucleation motifs, can be selectively stabilized *via* loop entropy perturbations, such as circular permutation. Over-stabilization of a nucleus may lead to frustration of the folding landscape, involving the segregation into local minima that compete for producing the native state. An example of such a scenario is represented by the D1pPDZ domain, a naturally occurring circularly permuted variant that displays an off-pathway kinetic trap characterized by a misfolded N-terminal hairpin incorrectly docked on an otherwise native-like structure (Gianni, et al. 2010).
Whilst the presentation of the funnel model provided a novel outlook on protein folding, a detailed experimental description of such a complex scenario is still a challenge.

1.8 Intrinsically disordered proteins and their binding and folding mechanisms

Contrary to the conventional view of the structure function dogma ‘sequence determines structure determines function’, another category of proteins have started to emerge in recent two decades called IDPs. Mostly IDPs come in the functionalan picture upon binding to their inteacting partners. The functions include the regulation of transcription and translation, cellular signalling, protein phosphorylation, the storage of small molecules and regulation of the self assembly of large multiprotein complexes such as bacterial flagellum and the ribosome (Dyson and Wright., 2005). Recent reviews have given the lights on their role as chaperones in plants and fungi (Tompa and Kovacs., 2010).

Many groups have reported that human genome contained sequences coding for unstructured proeins (Romero et. al. 1998). The outstanding statistics on the basis of bioinformatics and computational analysis revealed that 6-33% of bacterial proteins, 9-37% of archeal proteins and 35-50% of eukaryotic proteins should contain stretches of 40 or consecutive disordered amino acids (Dunker et. al. 2000). The experimental validation of this

statistical analysis is still to be done. On the same period many experimental biologists have reported that some proteins mainly of the transcriptional and translational control, cell cycle control and signalling remained without a stable 3D structure in all the ingenuous experimental conditions of buffer, salt and pH variations (Kriwacki et. al. 1996). Recently more focus has been given on IDPs which folds upon binding to its interacting structured partner (Dyson and Wright, 2005). Some IDPs get their folded form and eventually leading to function upon binding with another IDP partner (Dunker et. al. 2008). Sometimes metal binding acts as modifier of the IDP structure (Uversky et. al., 2000). Contrary to the fold and function paradigm there are some chaperones which get the functional form only upon getting unfolded. Activation of Hsp33 requires its native unfolding for function (Reichmann et al., 2012). Some other examples include the chaperones acid activated HdeA (Tapley et al., 2009) and the heat-activated small heat shock proteins (jaya et al., 2009), these chaperones use localized protein unfolding for activation.

Very briefly the introduction can be summarized in following ways:

1. Intrinsic disordered behaviour is encoded by its amino acid sequence: Like the ordered proteins the disordered proteins are also determined by their amino acid sequence. Generally a combination of low mean hydrophobicity and high net charge are the prerequisite for the absence of ordered structure in proteins (Uversky et. al. 2000). Much analysis revealed that IDPs were significantly depleted in so called order-promoting amino acids such as Ile, Leu, Val, Trp, Phe, Cys and Asn. The IDPs are substantially enriched in disorder promoting amino acids, Ala, Arg, Gly, Gln, Ser, Glu, Lys and Pro (Dunker et. al., 2001; Williams et. al., 2001; Romero et. al., 2001; Radivojak et. al., 2007).

2. Many IDPs follow the induced folding: many IDPs undergo a disorder to order transition upon interaction with specific binding partners (Dunker et. al., 2001; Dyson and Wright, 2002, 2005). Many IDPs bind very specifically to their ordered interacting partner while some IDPs can bind to many partners by changing the conformation to associate with different targets (Dunker et. al., 2001; Karush 1950; Kriwacki et. al. 1996). The C-terminal regulation domain of p53 which is disordered in its free form, adopts four different conformations , a short α-helix, a β-strand and two differently shaped coils, upon binding to its four interacting partners (Oldfield et. al. 2008). Contrary to the earlier case their are more than 10 IDPs which

bind with their ordered counterpart Kix domain (Dyson and Wright., 2005; Wang et. al. 2012).

3. Coupled folding and binding of IDPs: This is the process where an IDP or disordered region of a protein folds into an ordered structure concomitant with its target. There is an entropic cost to fold a disordered protein, which is paid for using the binding enthalpy. The most studied chapter is the pKID and Kix story. The Phosphorylated kinase-inducible domain (pKID) of transcription factor cyclic-AMP-response-element-binding protein (CREB) is unstructured when it is in free solution, but folds on forming a complex with the KID-binding (KIX) domain of CREB-binding protein (CBP) (Radhakrishnan et. al.1997).

4. Thermodynamics of coupled folding and binding: Binding of an IDP to its target is governed by the entropic cost associated with the disorder-to-order transition. Generally a favorable enthalpic contribution is the key driving force of a binding reaction, which gives an example of enthalpy-entropy compensation. Coupled folding and binding can give rise to various level of specificity and affinity depending on the subject. Generally in signal transduction and transcriptional IDPs make the complex with high specificity and low affinity so that after completing the process they must dissociate. Conformational flexibility of many IDPs facilitates the posttranscriptional modifications. pKID is the example where the specificity and affinity get a remarkable change upon phosphorylation. Several examples of entropy-enthalpy compensation are provided by the transcriptional activator CBP system e.g. pKID-KIX interaction, the hypoxia-inducible factor-1α (HIF1α), transcriptional-adapter-zinc-finger-1 (TAZ1) domain interaction and the activator for thyroid hormone and retinoid receptors (ACTR)-nuclear-receptor-co-activator-binding-domain (NCBD) interaction.

5. IDPs in diseases and drug target: on the basis of their biological functions and regulation in protein interaction networks many IDPs are implicated in various human diseases (Uversky et. al., 2008a; Midic et. al., 2009b). The c-Myc oncoprotein is a transcription factor that regulates large numbers of genes important in key cellular processes such as growth, differentiation, metabolism and apoptosis (Adhikary and Eilers, 2005). Its monomers are disordered and

undergo coupled folding and binding upon dimerization. Its overexpression occurs in most human cancers (Dang, 1999)

6. Computational prediction of IDPs: during the time of proteomics and progress of IDP research many predictors have been developed. Some of the most popular are PONDER, GlobPlot, DisEMBL, DISOPRED, FoldUnfold, IUPred, FoldIndex, DisPSMMP, IUP, POODLE-S, Bayes and CDF-ALL.

7. Biophysical Methods to characterize IDPs:
    a. Circular dichroism : far-UV CD spectroscopy gives the first basic information about the amount (or lack) of secondary structure. The ellipticity spectrum of IDPs has a large negative peak at nearly 200nm and values nearly close to zero at 220 nm. Partial structures in IDP can also be determined by these experiments.
    b. Nuclear magnetic resonance: heteronuclear magnetic resonance is the most powerful and versatile tool to see the disorderness in the protein. The high flexibility of the chain makes the chemical shift dispersion diminish and disorder easily recognizable (Dunker et al., 2001).
    c. There are many other techniques which are useful in this field are spectroscopic methods like Fourier transform infrared spectroscopy, provides additional information on the secondary structure. Near-UV CD, UV spectroscopy and fluorimetry experiments give the information about the environment of aromatic residues along with lack of a tightly packed hydrophobic core. The development of Raman optical activity measurement is paving the way for secondary structure determination (Syme et al., 2002).
    d. Some hydrodynamic techniques also give some information about the lack of structure. Gel filtration chromatography, small angle X-ray scattering, sedimentation analysis and dynamic light scattering provide information on hydrodynamic parameters. It includes stokes radius (R_s) and the radius of gyration (R_G).
    e. Other techniques including differential scanning calorimetry unveils the absence of cooperative folding transitions characteristic of a well defined tertiary fold. Many IDPs are heat stable so heat treatment also gives some useful information.

1.8.1 IDPs in transcriptional Activation: Partners of KIX domain of CBP

The IDPs which interact with the transcriptional coactivator CBP are the few earliest one to be studied in detail. Transcriptional coactivator CBP is a 2441 amino acid long multidomain protein, which serves as a central enabler in the transcription of, activated genes. It provides a bridge between the upstream activation site and the transcriptional initiation complex. One of the best-characterized interaction domains of CBP is the KIX domain, spanning residues 587-673. KIX domain is independently folded domain and follows a two state folding mechanism (Morrone et. al., 2012). KIX domain has been reported to bind as many as 10 IDPs discovered so far (Dyson 2011). Most of them are the transactivation domain of transcription factors e.g. c-myb, c-jun, tax, E2A-PBX1, MLL. pKId (kinase inducible domain) is also an IDP interacting with KIX.

1.8.2 Folding and binding measurements of Intrinsically disordered proteins with its interacting partner Kix

There are several hypotheses on the significance and value of disorder, which demand experimental validation (Uversky et al., 2010). From a thermodynamic viewpoint, it has been suggested that destabilization of a native fold may lower the affinity of a protein for its ligand, without necessarily compromising specificity (Wright and dyson, 1999; Spolar and record, 1994). Also, an interesting mechanistic model has been proposed by Wolynes and co-workers suggests the IDPs to display an increased capture radius to recruit and bind partners (Shoemaker et al., 2000). According to this view, called fly-casting mechanism, a disordered protein should form with its physiological partner a high-energy complex that would be locked in place by the coupled folding reaction. Whilst the kinetic data of binding and recognition of IDPs to their physiological partners are relatively scarce, it is of critical importance to understand the mechanisms whereby disordered systems operate (Keifhaber et. al., 2012).

The CREB-binding protein (CBP) is a co-activator that mediates the interaction between DNA-bound activator proteins and the components of the basal transcription complex. A small globular domain of CBP, known as KIX, modulates such interactions (Radhakrishnan et. al., 1997). Despite its small size, 87 amino acids, and a relatively simple fold, the KIX domain binds different IDP systems. Two different binding sites mediate such

interactions, namely the so-called c-Myb and MLL binding sites (named after two characteristic ligand of each of the site, i.e. the transactivation domain of the protein c-Myb and the mixed lineage leukaemia MLL protein) . Whilst the structural features of the interaction between KIX and different IDPs have been investigated in detail and several sets of co-ordinates are available in the pdb database (Radhakrishnan ey. Al., 1997; Zor et. al., 2004; Wang et. al. 2010), the mechanism of recognition is still a conundrum. More specifically, it is unclear whether KIX recognizes the IDPs in their ordered or disordered conformation – a critical question in tackling the proposition of an added value for disorder in IDP systems.

Objectives

The major objective of my PhD work was to characterize and understand the folding pathway of a bunch of heteromorphic protein pairs called Ga and Gb.

The problem of spontaneous folding of a polypeptide into compact, highly organized three-dimensional structures represents a fundamental challenge in modern Biochemistry (Kennedy and Norman 2005). The sequence of a protein contains all of the information for folding to the functionally competent state (Anfinsen 1961). However, not all amino acids of a polypeptide chain are equally important in specifying which fold to be adopted. The concept of "key residues" was previously proposed to describe the formation of "key contacts" in specific transition states of folding as a critical step to trigger "downhill" backbone collapse (Abkevich, et al. 1994; Vendruscolo, et al. 2001). The two proteins, called G_A88 and G_B88, with an extraordinarily high degree of sequence identity (i.e. 88%) but different structure and function (Alexander, et al. 2007) and also different folding mechanisms (Morrone et. al, 2011), offered to us a unique opportunity to start investigating the role of key residues in the mechanism of protein folding.

The surprising finding that the folding pathway of G_A88 and G_B88 diverges as early as in their denatured state (Morrone et. al, 2011), prompted us to carry out a systematic and very extensive analysis of each of the heteromorphic variants designed by Philip N. Bryan. All the pairs of G_A and G_B proteins produced by Bryan and co-workers, display structure and function similar to the respective wild-type proteins and interact specifically with two different macromolecules (i.e. serum albumin and IgG, respectively). These protein variants allowed us to explore the sequence space associated with a given protein structure. To this purpose, we performed a detailed characterization of the folding pathway of all six G_A and G_B proteins, i.e. G_A30, G_B30, G_A77, G_B77, G_A88 and G_B88, considering also the natural G_B domain, called GB1, a popular system for protein folding studies.
The results were obtained by exploring the folding process of GB1 at different pH values, and by performing an extensive Φ-value analysis of the engineered G_A and G_B variants. Analysis of data shows that the mechanism of folding of all G_A variants conforms to a co-operative two-state model with a

structurally conserved transition state. On the other hand, the folding pathway of the G_B variants appears less co-operative and more complex, revealing the presence of an intermediate and the existence of a variable organization in the transition states. Therefore, while the G_A-fold is populated via a single pathway with one conserved intrinsic nucleus, the G_B-fold can be reached by formation of more than one nucleus, each being selectively stabilized by altering the amino acid sequence via site-directed mutagenesis.

The last part of my Ph.D was focused towards understanding the folding and binding mechanism of cMyb, a member of the Intrinsically disordered proteins (IDPs) class. Trans Activation Domains (TADs) of cMyb is unfolded in physiological buffer condition is and only form the alpha helical structure upon binding to its interacting partner Kix domain (Radhakrishnan et. al. 1997, Dyson and Wright, 2005). In spite of the biological role of these IDPs, very little is known about the molecular details governing their binding and folding mechanism. Here the core aim was to deal with the molecular determinants of the coupled folding and binding mechanisms of the IDP, cMyb TAD and its interacting partner Kix.

METHODS

3.1 Site-directed Mutagenesis

G_A, G_B and Trans Activation Domain of c-Myb genes were cloned into the vector pG58, which encodes an engineered subtilisin pro-sequence as the N terminus of the fusion protein (Ruan, et al. 2004). These genes were used as templates to perform site-directed mutagenesis. All mutants were obtained by using the QuikChange mutagenesis kit (Stratagene) according to the manufacturer's instructions.

3.2 Protein Expression and Purification

Expression of the wild-type Ga-Gb proteins and their mutants was obtained in Luria Bertani (LB) medium containing 100μg/ml ampicillin. Cultures (1 liter in 2 liter flasks) were shaken at 180 rpm and grown at 37°C until OD600 arrived to 0,6; then protein expression was induced with 1 mM IPTG (isopropyl ß-d-thiogalactoside). After induction, cells containing Ga and Gb expression constructs were grown for 20 hours at 25°C, while c-Myb was grown for 20 hours at 37°C and then collected by centrifugation.

3.2.1 Purification of Ga and Gb along with its variants:

Cells containing expression constructs of Ga and Gb protein variants were resuspended in 100 mM $NaPO_4$ buffer (pH 7.2) and sonicated. After sonication, cell extract was centrifuged 30' at 13000 rpm to remove any insoluble material. Protein of interest was found in the soluble fraction.

Proteins were purified employing an affinity-chromatography previously developed (Ruan, et al. 2004). Soluble cell extract of pro-domain fusion proteins was injected on a 5-ml Bio-Scale™ Mini Profinity eXact cartridge at 5 ml/min to allow binding and then washed with 10-column volumes of 100 mM $NaPO_4$ (pH 7.2) to remove impurities. To cleave and elute the purified target proteins, 15 ml of 100 mM NaF in the presence of

100 mM $NaPO_4$ (pH 7.2) were injected at 5 ml/min. After the first 10 ml, the flow was stopped and the column incubated for 30 minutes to allow complete cleavage. The purified proteins were then dialyzed to remove sodium fluoride. Purity was checked by SDS/PAGE.

3.2.2 Purification of c-Myb Trans Activation Domain and its mutants:

Expression of c-Myb gene was done as described above. After sonication of the resuspended cells the protein was purified by both of its soluble and insoluble fraction. The column used was a cation-exchange cromatography (S-Sepharose equilibrated with 40 mM Tris-HCl pH 8.5). c-Myb* (c-Myb fused with pro-domain of subtilisin tag) was eluted with 850mM NaCl. The sample was then diluted 4-fold in 40 mM Tris, pH 8.5 and all minor impurities were removed by the purification step on a nickel(II)-charged chelating Sepharose FF (GE Healthcare) column equilibrated with 40 mM Tris-HCl equilibrated pH 8.5. The protein passed through the nickel(II)-column; the flow-through containing the protein was collected and concentrated. The purity of the protein was confirmed by SDS-PAGE.

3.2.3 Purification of Kix Domain:

WtKIX from Mus musculus CBP (residues 586-672) cloned into pRSET vector (Invitrogen) and resulting in a clone expressing His-tagged Kix, was generously provided by Per Jemth (Uppsala University, Sweden). The site-directed mutant Y658W named pwtKIX was obtained by using the QuikChange mutagenesis kit (Stratagene) according to the manufacturer's instructions. Cells containing expression constructs of Kix protein were resuspended in 40 mM Tris-HCl and 400 mM NaCl, pH 8.5 and sonicated. After sonication, cell extract was centrifuged 30' at 13000 rpm to remove any insoluble material. Proteins were purified by using a nickel(II)-charged chelating Sepharose FF (GE Healthcare) column equilibrated with 40 mM Tris-HCl and 400 mM NaCl, pH 8.5. The His-tagged wtKix and pwtKix were eluted with 250mM imidazole. The samples were then diluted 4-fold in 40 mM Tris, pH 8.5 and all minor impurities were removed by the purification step on a Q-Sepharose column equilibrated with 40 mM Tris, pH 8.5. The proteins passed through the Q-column; the flow-through containing the protein was collected and concentrated. Removal of hexa histidine tag was done with the standard procedure. The purity of the protein was confirmed by SDS-PAGE.

3.3 Equilibrium unfolding

Equilibrium denaturations were followed by Far-UV circular dichroism (CD) and fluorescence spectroscopy to monitor changes in secondary and tertiary structure, respectively.
CD measurements were carried out on a JASCO spectropolarimeter (Jasco, Inc., Easton, MD, USA), in a 1 cm quartz cuvette (Schellman). The spectra were recorded between 250 and 200 nm. Protein concentrations were typically 10 μM.
Fluorescence measurements were carried out on a Fluoromax single photon counting spectrofluorometer (Jobin-Yvon) with a cuvette of 1 cm light path. Tryptophan fluorescence emission spectra were recorded between 300 and 400 nm. The excitation wavelength was 280 nm. Protein concentrations were typically 1 to 6 μM.

3.4 Stopped-flow measurements

Single mixing kinetic folding and binding experiments were carried out on a Pi-star or on an SX-18 stopped-flow instruments (Applied Photophysics, Leatherhead, UK). The excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. In all folding experiments, performed at 25°C and 10°C, refolding and unfolding were initiated by a 11-fold dilution of the denatured or the native protein with the appropriate buffer. Final protein concentrations were typically 1 μM. The observed kinetics was always independent of protein concentration (from 0.5 to 5 μM), as expected from monomolecular reactions without effects due to transient aggregation (Silow and Oliveberg 1997).
The coupled folding and binding experiments of KIX and c-Myb were performed at 10°C by a two fold dilution of the binding partners. The protein concentration of KIX (or pwtKIX) was 10uM versus varying concentration of c-Myb.

3.5 Temperature-jump measurements

The observed kinetic rate constants of some G_A protein mutants were turning out to be beyond the range of stopped-flow apparatus. So they were obtained by Temperature-jump technique. The relaxation kinetics was measured as a function of guanidine or urea by using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (Hi-Tech, Salisbury, U. K.). Temperature was rapidly

changed from 18 °C to 25°C, 16 °C to 25°C and from 4 °C to 10 °C with a jump-size of 7°C, 9°C and 6°C respectively. 10 to 20 individual traces were averaged at given denaturant concentrations. The fluorescence change of N-acetyltryptophanamide (NATA) was used in control measurements. Protein concentrations of Ga variants were typically 20 μM. Binding and folding experiments on pwtKIX and c-Myb was performed with temperature jump from 16°C to 25°C just to compare with its stopped flow experiments. Protein concentration of pwtKIX was 10 μM versus varying concentration of c-Myb proteins. Degassed and filtered samples were slowly pumped through the 0.5 x 2 mm quartz flow cell before data acquisition. The excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter.

3.6 Buffers

Equilibrium and kinetic experiments were performed exploring a wide range of pH, from 2.0 to 10, using the following buffers: 50 mM Glycine /NaOH from pH 10 to 9.0, 50 mM Tris/HCl from pH 9.0 to 7.2, 50 mM sodium phosphate from pH 8.0 to 6.3, 50 mM Bis-Tris/HCl from pH 7.0 to 6.0, 50 mM sodium acetate from pH 5.5 to 3.8, 50 mM sodium formate from pH 3.4 to 3.0 and 50 mM sodium phosphate/phosphoric acid from pH 2.8 to 2.0. All reagents were of analytical grade.

3.7 Data analysis

3.7.1 Quantitative analysis of two-state equilibrium transitions

The folding-unfolding transition in globular proteins can be described, in general, as a two-state equilibrium process. Although there is now increasing evidence for complex behavior within such simple systems (Brockwell and Radford 2007), many small monomeric proteins show simple, unfolding transitions between native and denatured states. Under these conditions, if a given optical probe (y) is used to monitor unfolding, the observed signal y_{obs} will reflect the fractional population of the native and denatured state, being:

$$\frac{[D]}{[D]+[N]} = \frac{y_{obs} - y_N}{y_D - y_N} \qquad \text{(Eq. 3.1)}$$

$$\frac{[N]}{[D]+[N]} = 1 - \frac{y_{obs} - y_N}{y_D - y_N} \qquad \text{(Eq. 3.2)}$$

where y_{obs} is the optical measure at a given denaturant concentration, y_D and y_N are the signals of the denatured and native states respectively.
By applying the linear free energy extrapolation assumption, it can be postulated that:

$$\Delta G_{D-N} = \Delta G^0{}_{D-N} - m_{D-N}[denaturant] \quad \text{(Eq. 3.3)}$$

where ΔG_{D-N} is the stability of the protein at different denaturant concentrations and $\Delta G^0{}_{D-N}$ is the stability in water.
Moreover, according to the mass action law, for a two-state system:

$$K_{eq} = \frac{[D]}{[N]} \quad \text{(Eq. 3.4)}$$

and

$$\Delta G_{D-N} = -RT\ln(K_{eq}) \quad \text{(Eq. 3.5)}$$

where K_{eq} is the equilibrium constant of the reaction, ΔG_{D-N} is the stability of the native state to the denatured state, R is the gas constant and T is the temperature. Combining Eq. 3.1 with Eq. 3.4 and Eq. 3.5, the equation that defines the dependence of a given spectroscopic signal from the denaturant concentration can be derived:

$$y_{obs} = \frac{y_N + y_D e^{\left(\frac{(m_{D-N}[denaturant] - \Delta G^0{}_{D-N})}{RT}\right)}}{1 + e^{\left(\frac{(m_{D-N}[denaturant] - \Delta G^0{}_{D-N})}{RT}\right)}} \quad \text{(Eq. 3.6)}$$

In practice, this equation is complicated by the dependence of the intrinsic spectroscopic signal for both the native and the denatured state on the denaturant concentration (namely the native and denatured baselines (Santoro and Bolen 1988). Eq. 3.6, based on the two-state assumption, is now routinely used for the equilibrium studies of folding. A useful test to verify the two-state assumption is to calculate the thermodynamic folding parameters using different spectroscopic probes.

3.7.2 Quantitative analysis of folding kinetics: the two-state model

Analysis was performed by non-linear least-squares fitting of single exponential phases using the fitting procedures provided in the Applied Photophysics software. In a two-state reaction, as consequence of a perturbation induced on the system, the observed rate constant k_{obs} is represented by the following equation:

$$k_{obs} = k_F + k_U \qquad \text{(Eq. 3.7)}$$

where k_U and k_F represent the unfolding and refolding rate constants respectively. The logarithm of the observed (un)folding rate constants is linearly dependent on the denaturant concentration. Such a relationship is formalized in Eq. 3.8 (Jackson and Fersht 1991):

$$k_{obs} = k_F^{H_2O} e^{(-m_F[denaturant]/RT)} + k_U^{H_2O} e^{(m_U[denaturant]/RT)} \qquad \text{(Eq. 3.8)}$$

where k_F^{H2O} and k_U^{H2O} are the extrapolated folding and unfolding rate constants in absence of denaturant and, m_F and m_U reflect their dependence on denaturant concentration and correlate with the change in accessible surface area between the two ground states and the transition state between them.

Kinetic analysis of chevron plot allows the determination of the stability of a protein in the absence of denaturant agent ($\Delta G^0_{D\text{-}N}$):

$$\Delta G^0_{D-N} = -RT \ln\left(\frac{k_F}{k_U}\right) \qquad \text{(Eq. 3.9)}$$

where R is the gas constant and T the temperature.

Moreover, the algebric sum of the two kinetic *m*-values, m_U and m_F, allows to calculate the total $m_{D\text{-}N}$:

$$m_{D-N} = m_F + m_U \qquad \text{(Eq. 3.10)}$$

A comparison between the equilibrium and kinetic parameters $\Delta G^0_{D\text{-}N}$ and $m_{D\text{-}N}$ is crucial to verify the validity of the two-state model.

3.7.3 Quantitative analysis of folding kinetics: the three-state model

As reported in the Introduction Section,, if a partially folded intermediate is present, the folding kinetics can be described by the following scheme:

$$D \underset{k_{ID}}{\overset{k_{DI}}{\rightleftharpoons}} I \underset{k_{NI}}{\overset{k_{IN}}{\rightleftharpoons}} N \qquad \text{(Scheme 3.1)}$$

where k_{DI} is the microscopic rate constant for the formation of the intermediate from the denatured state, k_{ID} is the microscopic rate constant for the unfolding of the intermediate to the denatured state, k_{IN} is the microscopic rate constant for the formation of the native state from the intermediate and k_{NI} is the microscopic rate constant for the unfolding of the native state to the intermediate state.
Curved chevron plots were fitted, using the Kaleidagraph software package, by numerical analysis based on a three-state model following the equation:

$$k_{obs} = k_{DI} + \frac{k_{NI}}{++K_{part}} \qquad \text{Equ. 3.11}$$

where K_{part} is the partition factor k_{ID}/k_{IN} proportional to the difference between the activation barriers for the intermediate state to revert to the reagents rather than proceeding to the products. The logarithm of each microscopic rate constant was assumed to vary linearly with denaturant concentration (the slope of each dependence yielding the corresponding *m*-value).

3.7.4 Equilibrium binding analysis

Fluorescence experiments : Binding affinities of c-Myb to KIX were determined at 283 K by titrating 1 μM KIX (or pwtKIX) with the binding partner c-Myb (or c-Myb*), in a total volume of 1600 μl of 50 mM Sodium

Phosphate buffer, 150 mM KCl at pH 7.5. c-Myb concentration covered a suitable range from 0 to 100 μM.
The dissociation constant (K_D) was estimated using the Kaleidagraph software package. A nonlinear regression analysis was applied using hyperbolic equation $\Delta AU_{obs} = \Delta AU_{max}[S]/(K_D+[S])$ (Equation 3.1), where ΔAU_{obs} is the fluorescence difference ,ΔAU_{max} is the maximum fluorescence difference extrapolated to infinite ligand concentration, and [S] is the substrate analytical concentration.

RESULTS

4.1 Folding characterization of the different G_A and G_B heteromorphic pair variants

By taking the broad objective of deciphering the folding mechanisms of two protein variants with increasing degree of sequence identities, we performed a systematic analysis of each of the heteromorphic variants designed by Bryan and co-workers (Alexander, et al. 2007; He, et al. 2008), considering also the natural G_B domain, known as GB1. The proteins characterized with the folding perspective were: Gb1, Gb30, Gb77, Gb88, Ga30, Ga77, Ga88, KIX domain and c-Myb.

4.1.1 Folding characterization of GB1

To study the folding mechanism of GB1, both equilibrium and kinetic experiments were carried out. GB1 folding pathway was extensively characterized at 25 °C, under different pH conditions, exploring a wide range from 2.0 to 9.6.

4.1.1.1 Equilibrium denaturations

Guanidine-induced equilibrium denaturations of GB1, monitored by fluorescence spectroscopy, were obtained at 25°C exploring a wide range of pH, from 2.0 to 9.6 (Figure 4.1). Equilibrium denaturations were fitted both individually and globally with shared $m_{D\text{-}N}$ value. Values obtained from the global analysis are listed in Table 4.1. They were consistent within error with the values obtained by fitting individually each independent equilibrium experiments, as well as with the value calculated from kinetic experiments, confirming the two-state nature of the equilibrium unfolding transition of GB1. The denaturation profiles were consistent with a two-state unfolding and returned an *m*-value of 1.75 ± 0.2 kcal mol^{-1} M^{-1}.

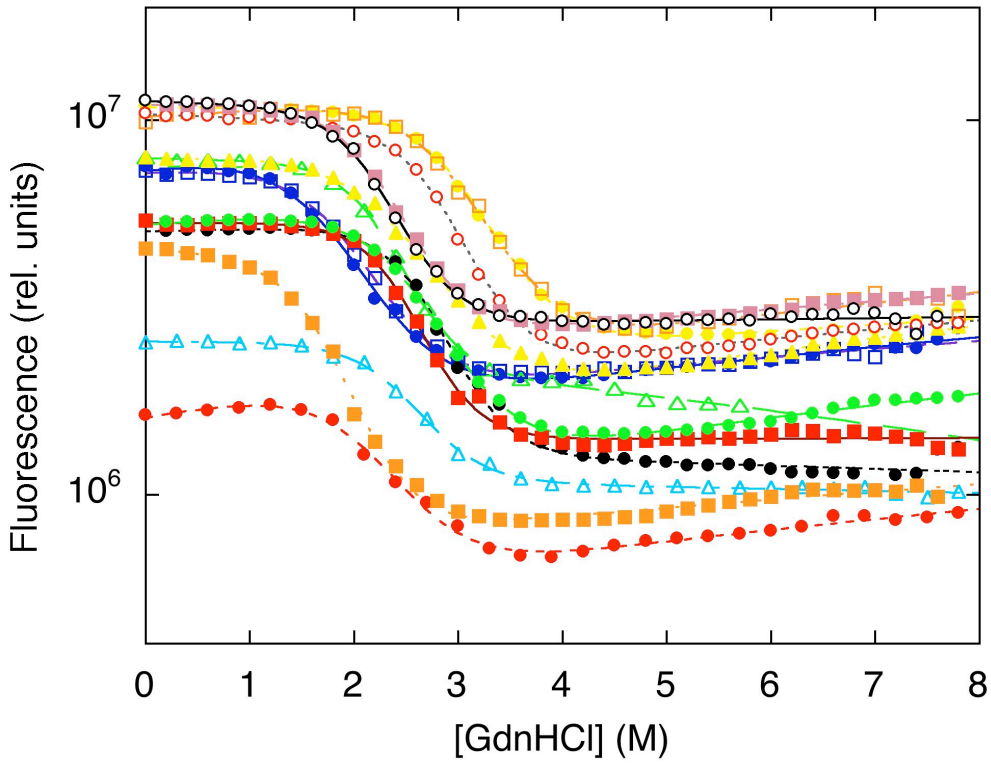


Figure 4.1. Equilibrium denaturations of GB1 monitored by fluorescence in a wide range of pH, from 2.0 to 9.6, at 25°C (●, pH 2.0; △, pH 2.5; □, pH 3.0; ●, pH 3.5; ▲, pH 4.0; △, pH 4.5; ●, pH 5.0; □, pH 6.0; ●, pH 6.5; ○, pH 7.0; ■, pH 7.5; ●, pH 8.0; ■, pH 8.5; ○, pH 9.0; ■, pH 9.6). Lines represents the best fit to a two state unfolding transition.

4.1.1.2 Kinetic experiments

The folding and unfolding kinetics of GB1 were investigated at several pH values, ranging from 2.0 to 9.6. In all cases, folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final denaturant concentration. Semi-logarithmic plots of the observed folding/unfolding rate constants of GB1 versus guanidine concentration (chevron plots) at the different pH values are presented in Figure 4.2.

Surprisingly, the unfolding arm of the chevron plots at pH values higher than 6.0 shows a deviation from linearity that becomes evident at high guanidine concentrations (roll over effect). This effect escaped previous studies probably because of the restricted

range of experimental conditions, limited to [GdnHCl] < 5.5 M (Park, et al. 1997; McCallister, et al. 2000).
Indeed, if we were to ignore the data we recorded for [GdnHCl] > 5.5 M, the unfolding arm of the chevron plots would appear essentially linear but would display a puzzling change in slope in the different experimental conditions.

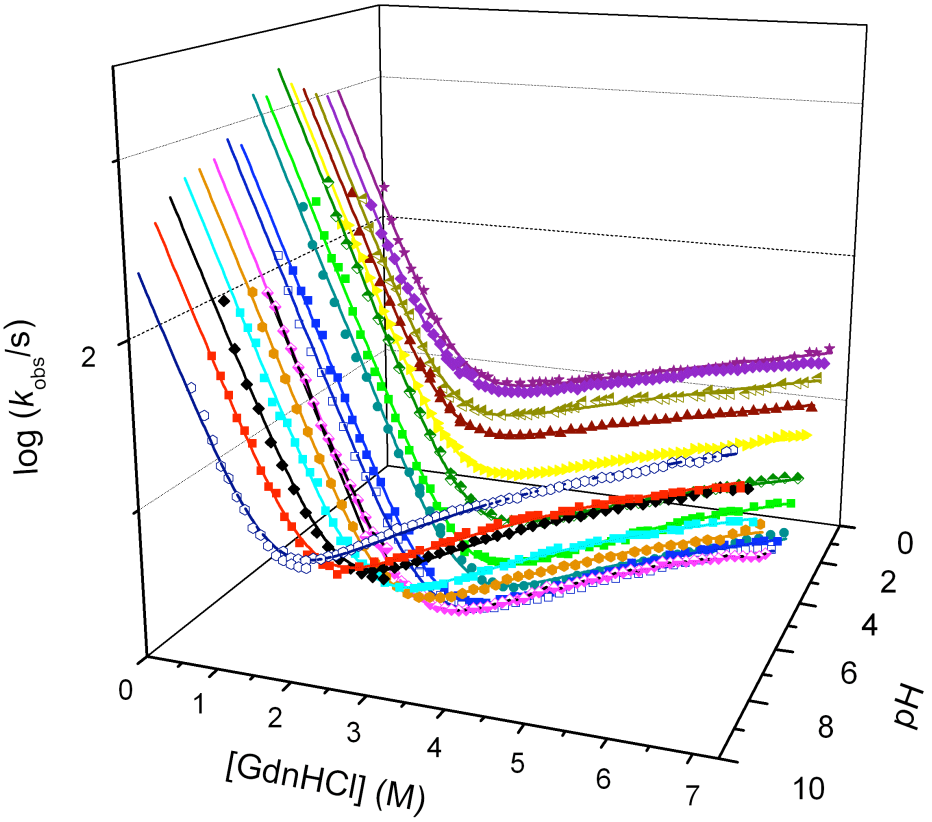


Figure 4.2. Chevron plots of GB1 measured from pH 2.0 to pH 9.6. Three-dimensional graph was obtained with Origin software. Lines are the best global fit to a three-state equation with shared m-values (Gianni, et al. 2007b). Exclusion of data [GdnHCl] > 5.5 M would result in a quasi-linear unfolding arm with an apparent change in unfolding m-values.

The deviation from linearity of the chevron plots of GB1is highlighted in Figure 4.3 where the chevron plot obtained at pH 9.0 is reported together with the residuals of the

fit, showing a clear systematic deviation from the expected values for a two-state behavior.
As reported in the Introduction Section, a deviation from linearity in either the folding or the unfolding branches may be considered of diagnostic value for the identification of intermediates (Parker, et al. 1995; Wildegger and Kiefhaber 1997). If a partially folded intermediate is present, the folding kinetics can be described by a three-state mechanism and the observed chevron plots can be fitted by numerical analysis based on a three-state model following the Equation 3.11 reported in Methods Section.

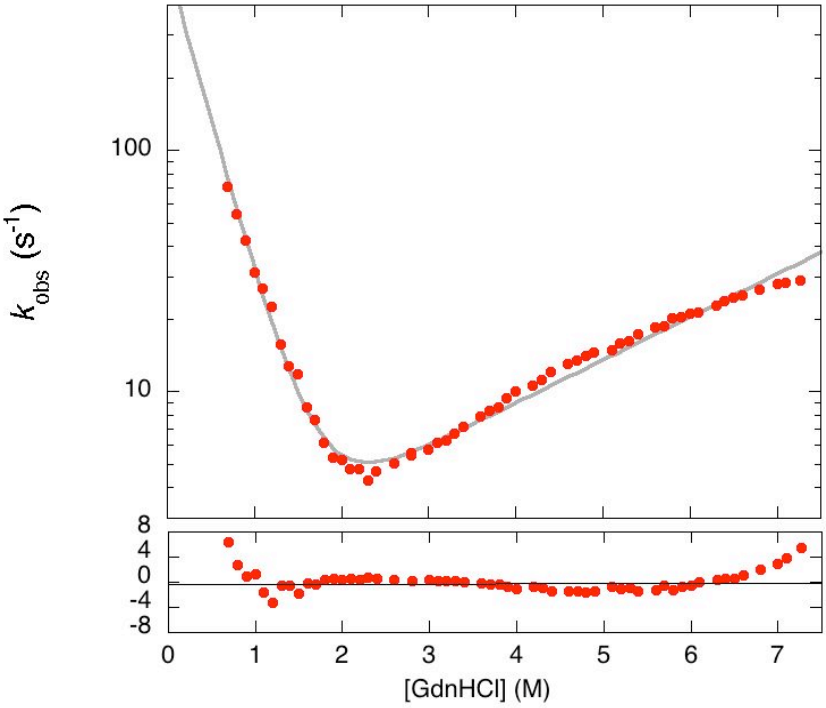


Figure 4.3. Chevron plot of wild-type GB1 measured at pH 9.0. The grey line is the best fit to a two-state equation. The residuals of the fit, showing a clear systematic deviation from the expected values for a two-state behavior, are reported below the graph. Inclusion of data at [GdnHCl] > 5.5 M are critical to detect the roll-over effect.

Parameters calculated from global analysis (listed in Table 4.1) allow the identification of the relative positions of the two activation barriers along the reaction coordinate in terms of their relative accessible surface area (Tanford β-value), resulting in a β_T-value of 0.76 ± 0.04 for the transition state *TS1* and 0.93 ± 0.04 for the native-like activation

barrier *TS2*. The excellent statistical parameters of the global analysis indicate that the two activation barriers are robust to changes in pH conditions and display a conserved solvent accessible surface area when pH and protein stability are altered.

pH	k_{DI} (s^{-1})	$k_{NI}/(1+K_{part})$ (s^{-1})	K_{part}	$\Delta G_{D\text{-}N}$ [a] (kcal mol^{-1})	$\Delta G_{D\text{-}N}$ [b] (kcal mol^{-1})
2.0	1000 ± 90	11.0 ± 3.0	1.7 ± 0.2	2.7 ± 0.3	3.9 ± 0.8
2.5	1000 ± 90	19.2 ± 4.8	3.0 ± 0.4	2.3 ± 0.2	3.0 ± 0.7
3.0	1050 ± 95	5.2 ±1.3	0.9 ±0.1	3.1 ± 0.4	3.8 ± 0.4
3.5	1200 ± 120	5.1 ±1.3	1.2 ±0.1	3.2 ± 0.2	3.9 ± 0.4
4.0	1400 ± 120	1.3 ± 0.3	0.42 ± 0.09	4.1± 0.2	4.5 ± 0.4
4.5	1800 ± 170	0.4 ± 0.1	0.20 ± 0.04	5.0 ± 0.2	4.2 ± 0.4
5.0	1300 ± 120	0.15 ± 0.04	0.09 ± 0.02	5.3 ± 0.2	4.9 ± 0.5
5.5	1370 ± 100	0.15 ± 0.04	0.13 ± 0.03	5.4 ± 0.1	4.6 ± 0.5
6.0	720 ± 70	0.12 ± 0.03	0.09 ± 0.02	5.1 ± 0.1	5.2 ± 0.5
6.5	830 ± 70	0.14 ± 0.04	0.10 ± 0.02	5.1 ± 0.2	4.7 ± 0.5
7.0	670 ± 70	0.19 ± 0.05	0.12 ± 0.02	4.8 ± 0.2	5.3 ± 0.5
7.5	630 ± 70	0.27 ± 0.07	0.11 ± 0.02	4.6 ± 0.2	4.3 ± 0.4
8.0	600 ± 60	0.4 ± 0.1	0.12 ± 0.03	4.3 ± 0.3	4.3 ± 0.4
8.5	500 ± 40	0.6 ± 0.2	0.11 ± 0.02	3.9 ± 0.2	3.8 ± 0.4
9.0	390 ± 40	0.9 ± 0.2	0.13 ± 0.03	3.6 ± 0.2	3.3 ± 0.3
9.6	220 ± 25	1.3 ± 0.3	0.11 ± 0.02	3.0 ± 0.2	3.0 ± 0.3

Table 4.1. Folding parameters of GB1 as a function of pH. [a]Calculated from chevron plot analysis. The chevron plots were fitted globally to a three-state model with shared m values. k_{DI} is the microscopic rate constant for the formation of the intermediate from the denatured state; k_{NI} is the microscopic rate constant for the unfolding of the native state to the intermediate state; K_{part} is the partitioning factor k_{ID}/k_{IN} reflecting the difference between the activation barriers for the intermediate to revert to the reagents rather than proceeding to the products. The analysis returned a total $m_{D\text{-}N}$ = 1.95 ± 0.2 kcal mol^{-1} M^{-1}. The Tanford β values for the two transition states were β_{TS1} = 0.76 ± 0.04 and β_{TS2} = 0.93 ± 0.04. [b]Calculated from equilibrium experiments. Equilbrium denaturation curves were fitted both individually and globally with shared $m_{D\text{-}N}$ value. The global analysis returned $m_{D\text{-}N}$ = 1.75 ± 0.2 kcal mol^{-1} M^{-1}. This value was consistent within error with the values obtained by fitting individually each independent equilibrium experiments, as well as with the value calculated from kinetic experiments.

4.1.1.3 The effect of pH on the folding kinetics of GB1

The analysis of the chevron plots reported in Figure 4.8 allowed measuring the folding and unfolding rate constants of GB1 over a very wide range of pH. The folding rate constants display a negligible dependence on pH (data not shown). A plot of the logarithm of apparent unfolding rate constants from the native state to *TS2* (k_{NI}) and *TS1* (formally equivalent to $k_{NI}/(1+K_{part})$) as a function of pH are reported in Figure 4.10.

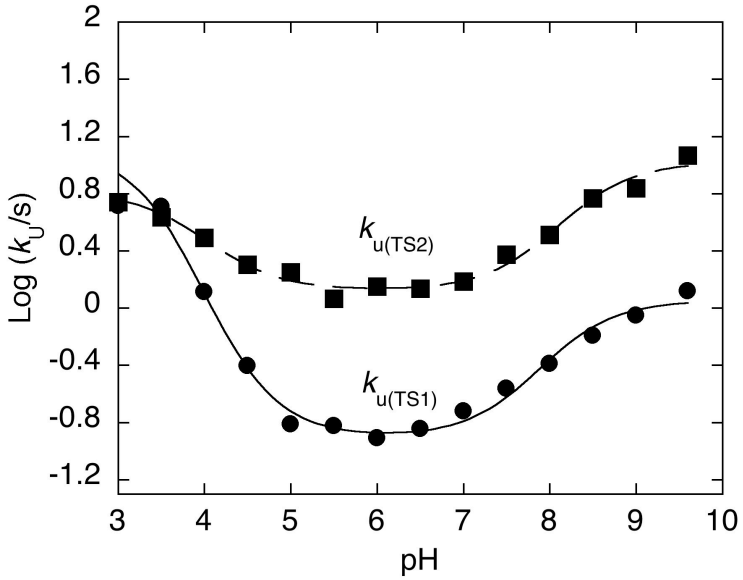


Figure 4.4 Unfolding rate constants versus pH. Logarithm of calculated unfolding rate constants from the native state N to the first transition state k_u(TS1) and from the native state N to the second transition state k_u(TS2) as a function of pH. The lines are the best fit to a model involving the protonation of two groups. In both cases, we obtained the same approximate pK_a of ~ 4 and ~ 8.

Interestingly, both *TS1* and *TS2* display sigmoidal transitions at acidic and alkaline pH consistent with protonation of at least two groups in the native state with apparent pK_a of ~ 4 and ~ 8. Importantly, the acid transition for the unfolding rate constant of *TS1* (changing by almost two orders of magnitude) is more pronounced than that for the unfolding rate constant of *TS2* (changing by less than an order of magnitude),

suggesting the contribution of a salt bridge that is weak or not formed in *TS1*, but is consolidated in *TS2*.

4.1.2 Φ-value analysis of the G_A and G_B variants

To test how sequence composition versus topology dictates the folding of proteins, we performed an extensive Φ-value analysis on six different proteins (see Figure 1.9 in introduction section), three variants of the G_A domain of staphylococcal protein G (i.e. G_A30, G_A77 and G_A88), and three variants of the G_B domain (i.e. G_B30, G_B77 and G_B88). A total of 132 mutants were produced, purified and characterized by equilibrium and kinetic folding experiments.

Performing a complete Φ-value analysis demands a careful selection of the experimental conditions. In fact, the protein of interest must be stable enough to allow accurate determination of the folding kinetics for its destabilized mutants, but it must not be too stable otherwise its unfolding kinetics may be difficult to evaluate with the necessary accuracy. As detailed below, since the three pairs of proteins characterized in this work and their site directed variants displayed widely different thermodynamic stabilities as well as different sensitivities to changes in ionic strength, we optimized the experimental conditions for each protein with regard to temperature and type of denaturant (i.e. urea or GdnHCl). In addition, in an effort to test the robustness of the folding pathway, a limited set of Φ values was measured for each protein at more than one experimental condition, while the full set of Φ values was obtained for each system at pH 7.2.

G_A proteins. Urea induced equilibrium transitions of G_A30, G_A77 and G_A88 were measured at 25°C and pH 7.2 in 50 mM sodium phosphate buffer. A typical equilibrium denaturation profile for each protein is reported in Figure 4.5 (A-C).

While in the case of G_A77 and G_A88 we could observe monotonic sigmoidal transitions, G_A30 was found to be too stable and was not fully denatured even at very high urea concentrations; therefore, in the case of G_A30 we performed (un)folding experiments using GdnHCl. Because of the ionic nature of GdnHCl, we could not perform (un)folding experiments with this denaturant on G_A77 and G_A88, since their stabilities display a pronounced dependence on ionic strength, as shown by folding experiments in the presence of sodium chloride.

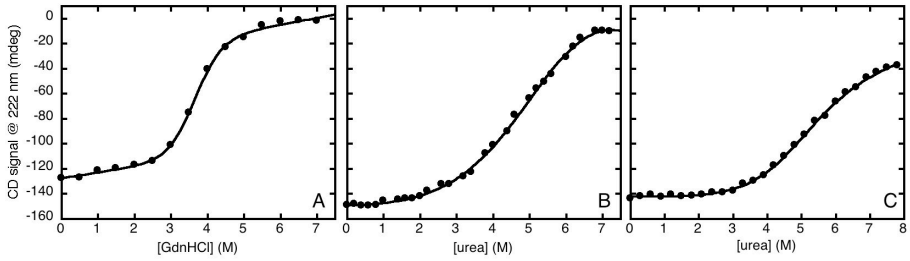


Figure 4.5. Equilibrium denaturation of G_A (up panels) variants monitored by CD in 50mM sodium phosphate buffer at pH 7.2. As detailed in the text, because of the very different thermodynamic stabilities of the proteins, the equilibrium denaturations were performed at different experimental conditions: the unfolding experiment of G_A30 was performed at 25°C using GdnHCl; G_A77 and G_A88 experiments were performed at 10°C using urea.

We extensively studied the folding and unfolding kinetics of G_A30, G_A77 and G_A88 by stopped-flow and temperature jump (T-jump) experiments. In the case of G_A77 and G_A88, it was not possible to measure reliable folding and unfolding rate constants at 25 °C over a wide range of denaturant concentration, because reactions were too fast for our stopped-flow apparatus; thus, kinetic folding data for the two proteins were recorded at 10 °C. In all cases, the folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final denaturant concentration. Semi-logarithmic plots of the observed folding/unfolding rate constants of G_A30, G_A77 and G_A88 versus denaturant concentration (i.e. chevron plots) are presented in Fig. 4.6 (A-C). All proteins displayed a V-shaped chevron plot, a hallmark of two-state folding (Fersht, 1999). Two-state folding was further confirmed by the excellent agreement between the thermodynamic parameters obtained by equilibrium and kinetic data. These data parallel earlier studies on G_A88 (Morrone et. a., 2011), and confirm that this protein system folds *via* a two-state folding pathway with an unstructured denatured state.

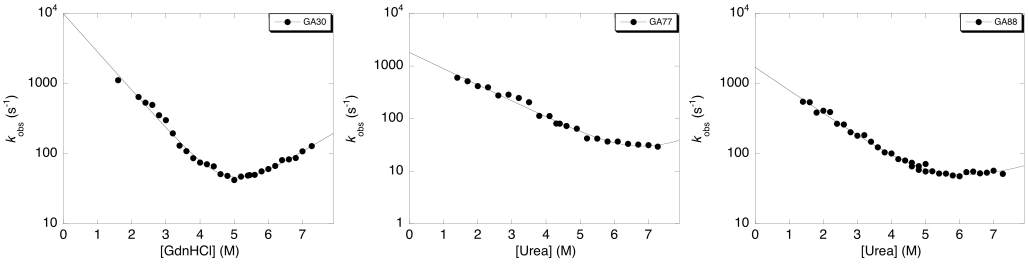


Figure 4.6. Chevron plots of G_A variants. Semilogarithmic plot of the observed rate constant for folding and unfolding of G_A30, G_A77 and G_A88 versus [denaturant] measured at pH 7.2.

We addressed the structural features of the transition state for folding of the G_A proteins by Φ-value analysis. A total of 50 mutants were produced: 14 for G_A30, 17 for G_A77 and 19 for G_A88. Three mutants expressed poorly or were too unstable to be included in the analysis. The remaining 47 were subjected to equilibrium and kinetic folding experiments (Figures 4.7, 4.8, and 4.9).
In some cases, the folding and unfolding rate constants were too fast for the stopped-flow methodology and were determined using a capacitor-discharge T-jump apparatus. Fitted parameters are listed in Table 4.2, 4.3 and 4.4.

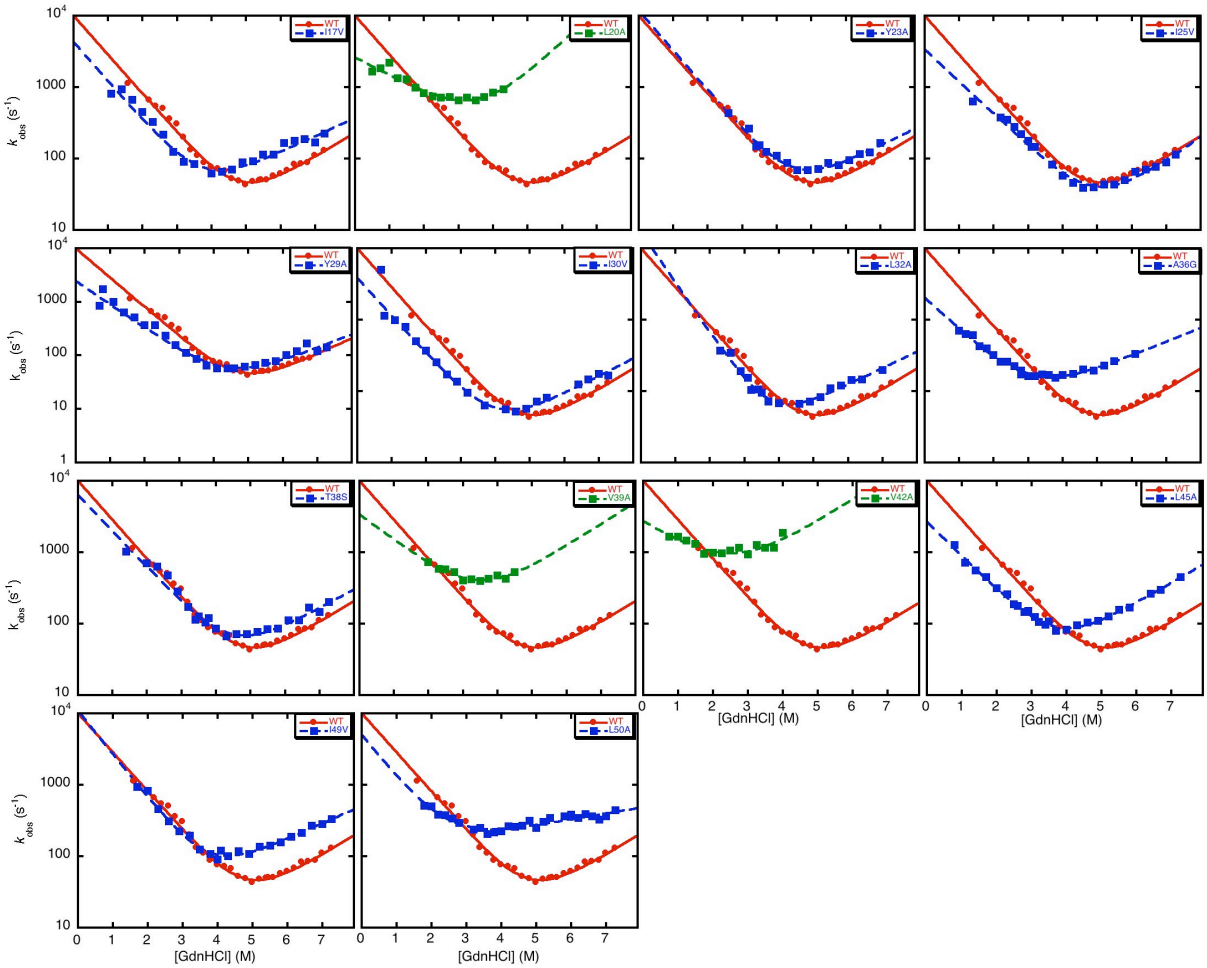


Figure 4.7. Chevron plots of wild-type (red filled circles) and mutants (blue and green filled squares) of G_A30. Observed rate constants were measured as a function of GdnHCl at pH 7.2 and 25°C. All the chevron plots were obtained by employing the stopped-flow apparatus, except for the mutants L20A, V39A and V42A (green filled squares) measured by Temperature-jump. Lines are the best fit to a two-state model.

Table 4.2. Kinetic and thermodynamic parameters for the folding of G_A30 site-directed mutants, at pH 7.2 and 283 K.

	k_F (s^{-1})	k_U (s^{-1})	ΔG_{D-N} (kcal mol^{-1})	$\Delta\Delta G_{D-TS}$ (kcal mol^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})	Φ	Tm (K)
WT	7800 ± 250	2.4 ± 0.1	4.5 ± 0.02				357 ± 9
I17V	4400 ± 130	5.1 ± 0.2	3.8 ± 0.02	0.32 ± 0.05	0.75 ± 0.01	0.43 ± 0.06	357 ± 10
L20A	5900 ± 100	98 ± 2	2.3 ± 0.01	0.16 ± 0.03	2.2 ± 0.01	0.07 ± 0.01	341 ± 2
Y23A	8500 ± 350	3.7 ± 0.2	4.3 ± 0.02	0.05 ± 0.04	0.19 ± 0.01	*	361 ± 25
I25V	5900 ± 300	2.4 ± 0.2	4.4 ± 0.03	0.15 ± 0.05	0.17 ± 0.01	*	360 ± 2
Y29A	4480 ± 100	3.4 ± 0.1	4.0 ± 0.01	0.31 ± 0.04	0.51 ± 0.01	0.61 ± 0.08	353 ± 6
I30V	3800 ± 100	4.2 ± 0.1	3.8 ± 0.01	0.41 ± 0.05	0.72 ± 0.01	0.56 ± 0.06	359 ± 22
L32A	5300 ± 200	4.7 ± 0.2	3.9 ± 0.02	0.22 ± 0.04	0.60 ± 0.01	0.37 ± 0.07	348 ± 1
A36G	3300 ± 100	16 ± 0.4	3.0 ± 0.01	0.48 ± 0.05	1.5 ± 0.01	0.31 ± 0.03	345 ± 2
T38S	8200 ± 250	3.8 ± 0.2	4.3 ± 0.02	0.03 ± 0.03	0.24 ± 0.01	*	357 ± 3
V39A	7700 ± 100	45 ± 1	2.9 ± 0.01	0.01 ± 0.02	1.7 ± 0.02	0.00 ± 0.01	344 ± 5
V42A	3600 ± 100	210 ± 10	1.6 ± 0.01	0.43 ± 0.05	2.9 ± 0.02	0.15 ± 0.02	342 ± 2
L45A	3420 ± 100	7.6 ± 0.2	3.4 ± 0.01	0.46 ± 0.05	1.1 ± 0.02	0.42 ± 0.04	354 ± 6
I49V	6700 ± 120	7.0 ± 0.2	3.8 ± 0.01	0.08 ± 0.03	0.69 ± 0.01	0.12 ± 0.04	356 ± 10
L50A	5100 ± 110	21 ± 1	3.1 ± 0.01	0.24 ± 0.04	1.5 ± 0.01	0.16 ± 0.03	345 ± 2

*Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_A30 proteins calculated from chevron plot analysis and equilibrium heat-induced denaturation. The chevron plots were fitted globally to a two-state model. The Tanford β-value for the transition state was $\beta_{TS} = 0.69 \pm 0.04$. * Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate a reliable Φ value.*

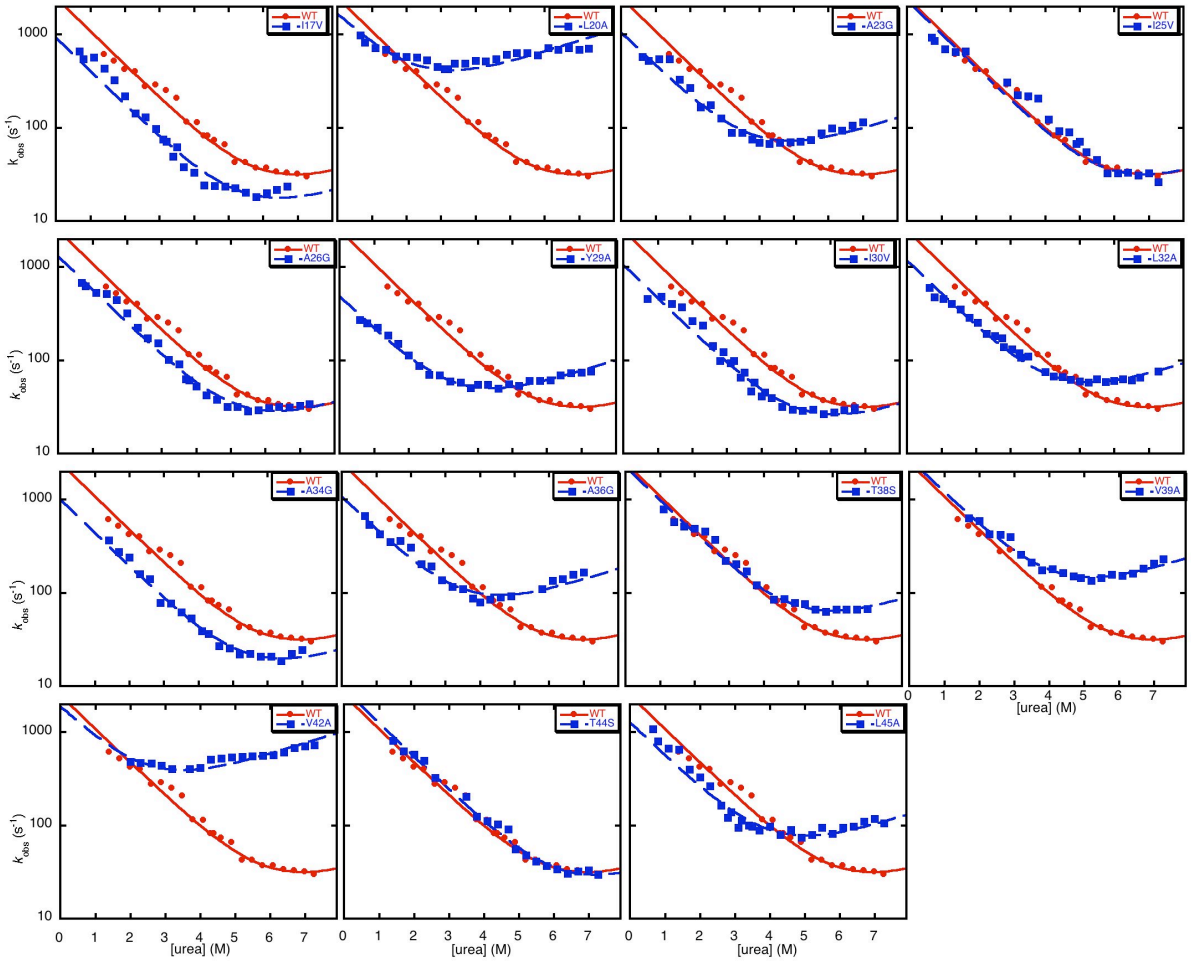


Figure 4.8. Chevron plots of wild-type (red filled circles) and mutants (blue filled squares) of G_A77. Observed rate constants were measured as a function of urea at pH 7.2 and 10°C. All the chevron plots were obtained by employing the stopped-flow apparatus. Lines are the best fit to a two-state model.

Table 4.3. Kinetic and thermodynamic parameters for the folding of G_A77 site-directed mutants, at pH 7.2 and 283 K.

	k_F (s^{-1})	k_U (s^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})	Φ	Tm (K)
WT	2460 ± 50	3.8 ± 0.2			350 ± 2
A12G	*				
A16G	*				
I17V	990 ± 20	2.4 ± 0.1	0.26 ± 0.01	†	337 ± 26
L20A	1540 ± 30	136 ± 10	2.3 ± 0.02	0.12 ± 0.01	352 ± 22
A23G	1100 ± 60	14 ± 1	1.2 ± 0.01	0.37 ± 0.03	331 ± 1
I25V	2870 ± 100	3.2 ± 0.3	0.17 ± 0.01	†	328 ± 1
A26G	1320 ± 70	4.0 ± 0.1	0.38 ± 0.01	0.91 ± 0.07	339 ± 2
Y29A	480 ± 30	12 ± 1	1.5 ± 0.01	0.59 ± 0.04	352 ± 6
I30V	1120 ± 50	3.8 ± 0.2	0.45 ± 0.01	0.98 ± 0.07	343 ± 5
L32A	1200 ± 50	11 ± 0.3	0.98 ± 0.01	0.41 0.03	335 ± 5
A34G	990 ± 50	2.9 ± 0.1	0.36 ± 0.01	1.40 ± 0.09	341 ± 1
A36G	1100 ± 50	21 ± 0.6	1.4 ± 0.01	0.32 ± 0.02	341 ± 2
T38S	2300 ± 80	10 ± 0.3	0.56 ± 0.02	0.07 ± 0.03	337 ± 3
V39A	2800 ± 100	27 ± 2	1.0 ± 0.04	0.07 ± 0.01	325 ± 1
V42A	1600 ± 40	119 ± 11	2.2 ± 0.01	0.11 ± 0.01	343 ± 4
T44S	3000 ± 50	3.1 ± 0.1	0.22 ± 0.01	†	321 ± 2
L45A	1370 ± 50	15 ± 0.6	1.1 ± 0.01	0.30 ± 0.03	317 ± 5

*Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_A77 proteins calculated from chevron plot analysis and equilibrium heat-induced denaturation. The chevron plots were fitted globally to a two-state model. The Tanford β-value for the transition state was $\beta_{TS} = 0.75 \pm 0.04$. *These mutants expressed poorly and could not be characterized. † Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate a reliable Φ value.*

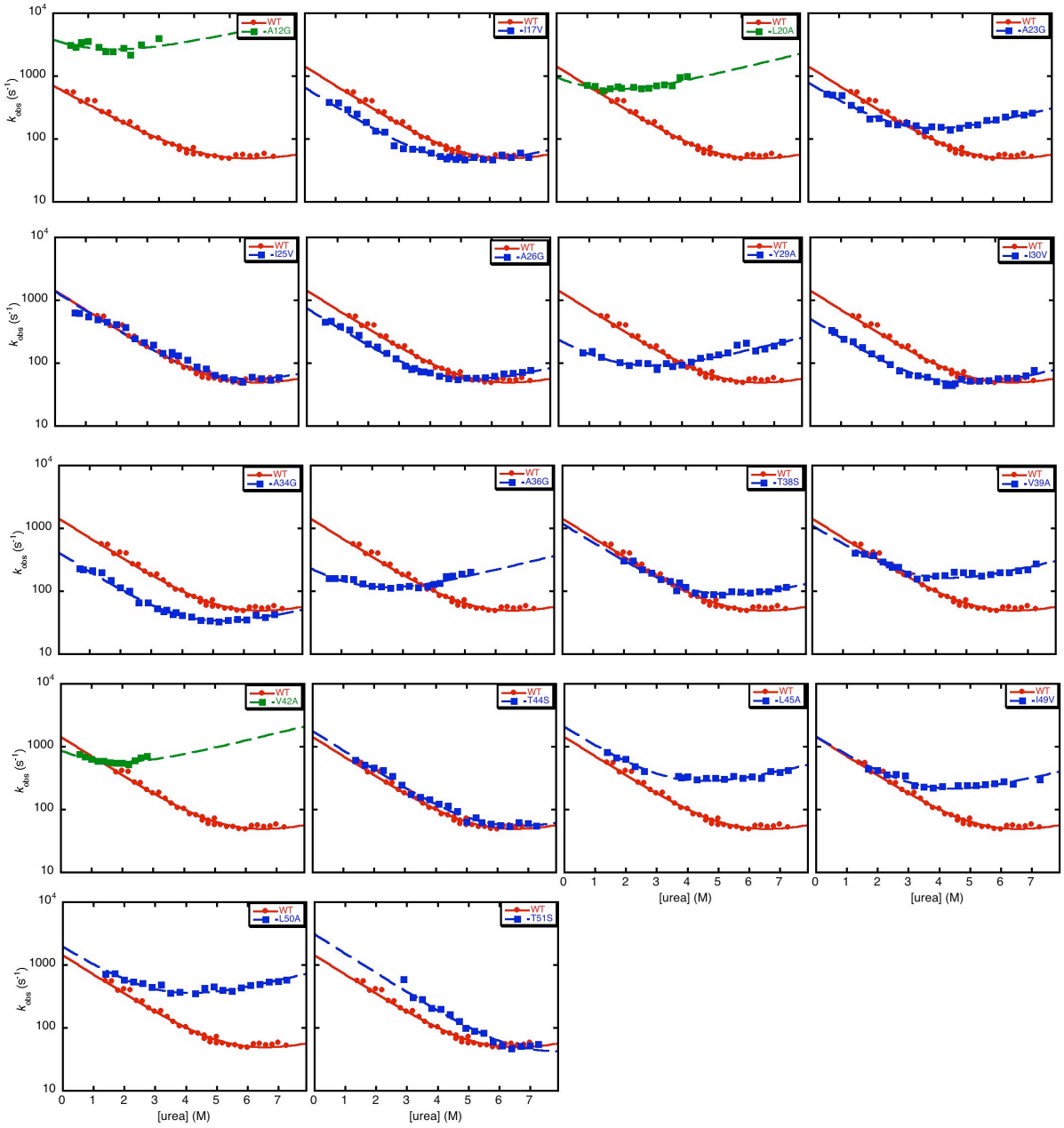


Figure 4.9. Chevron plots of wild-type (red filled circles) and mutants (blue and green filled squares) of G_A88. Observed rate constants were measured as a function of urea at pH 7.2 and 10°C. All the chevron plots were obtained by employing the stopped-flow

apparatus, except for the mutants A12G, L20A and V42A (green filled squares) measured by Temperature-jump. Lines are the best fit to a two-state model.

Table 4.4. Kinetic and thermodynamic parameters for the folding of G_A88 site-directed mutants, at pH 7.2 and 298 K.

	k_F (s^{-1})	k_U (s^{-1})	$\Delta\Delta G_{D-N}$ (kcal mol^{-1})	Φ	Tm (K)
WT	1500 ± 100	6.4 ± 0.8			345 ± 4
A12G	5000 ± 780	1000 ± 90	2.1 ± 0.02	0.32 ± 0.00	346 ± 2
A16G	*				
I17V	700 ± 30	7.7 ± 0.8	0.52 ± 0.01	0.81 ± 0.14	333 ± 3
L20A	670 ± 110	280 ± 20	2.5 ± 0.02	0.18 ± 0.00	333 ± 4
A23G	770 ± 40	37 ± 3	1.3 ± 0.02	0.27 ± 0.06	336 ± 5
I25V	1500 ± 30	7.6 ± 0.9	0.09 ± 0.00	†	329 ± 6
A26G	760 ± 30	10 ± 1.0	0.62 ± 0.01	0.61 ± 0.11	340 ± 2
Y29A	210 ± 20	32 ± 3	2.0 ± 0.09	0.55 ± 0.11	340 ± 1
I30V	500 ± 20	9.4 ± 0.9	0.81 ± 0.01	0.74 ± 0.12	320 ± 3
A34G	410 ± 3	6.1 ± 0.1	0.68 ± 0.01	1.04 ± 0.14	336 ± 3
A36G	200 ± 4	42 ± 1	2.2 ± 0.03	0.52 ± 0.09	338 ± 3
T38S	1200 ± 20	16 ± 0.3	0.61 ± 0.01	0.20 ± 0.06	333 ± 4
V39A	950 ± 20	39 ± 1	1.2 ± 0.01	0.20 ± 0.04	334 ± 3
V42A	610 ± 10	260 ± 25	2.6 ± 0.06	0.20 ± 0.03	354 ± 6
T44S	1770 ± 60	7.1 ± 0.5	0.05 ± 0.01	†	322 ± 3
L45A	1930 ± 80	63 ± 7	1.9 ± 0.02	0.07 ± 0.01	341 ± 3
I49A	1210 ± 40	51 ± 5	1.3 ± 0.01	0.09 ± 0.03	328 ± 1
L50A	1970 ± 30	89 ± 8	1.3 ± 0.01	0.12 ± 0.02	339 ± 2
T51S	3050 ± 50	4.3 ± 0.1	0.63 ± 0.01	0.64 ± 0.03	329 ± 3

*Kinetic and thermodynamic parameters of wild-type (WT) and mutant G_A88 proteins calculated from chevron plot analysis and equilibrium heat-induced denaturation. The chevron plots were fitted globally to a two-state model. The Tanford β-value for the transition state was $\beta_{TS} = 0.72 \pm 0.04$. *These mutants expressed poorly and could not be characterized. † Mutants with $\Delta\Delta G_{D-N}$ too low (<0.35 kcal mol^{-1}) to calculate a reliable Φ value.*

Following a generally accepted convention (Fersht et. al., 1992; Geierhaas et. al., 2008; Gianni et. al., 2010; Ivarsson et. al., 2009), the experimentally determined Φ-values were grouped in three different classes and mapped on the native structure of the G_A protein (Figure 4.10): small values (Φ<0.3; red), intermediate values (0.3 < Φ < 0.7; magenta) and large values (Φ>0.7; blue). Inspection of Figure 4.10 clearly reveals that the folding nucleus is by-and-large conserved in all the G_A proteins, with the highest Φ values clustered at the interface between helix 1 and helix 2, suggesting that, in the case of such a two-state system, the dominant folding mechanism is very robust to perturbations of the primary structure.

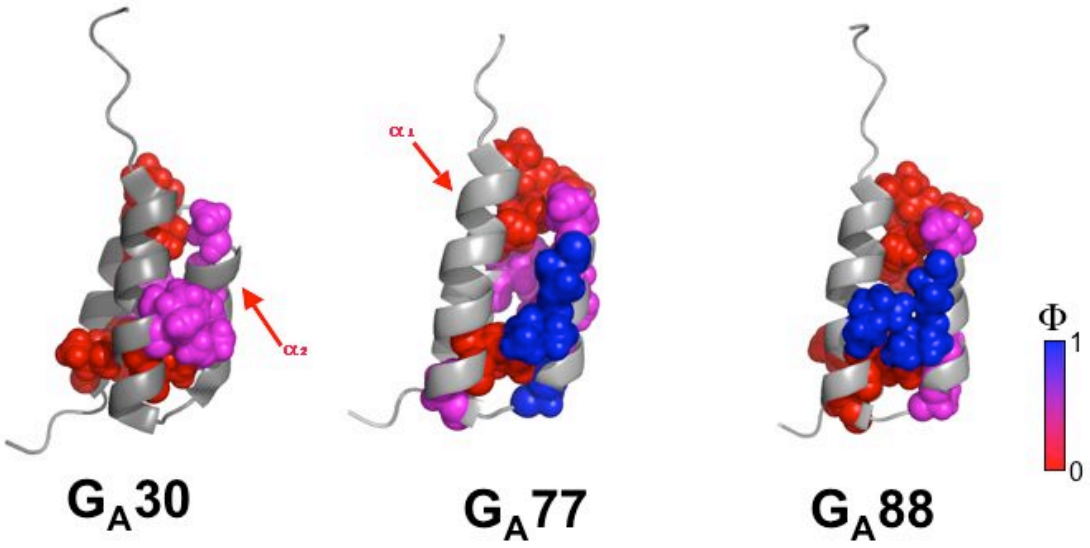


Figure 4.10. Structural distribution of the measured Φ-values on the native structures of G_A30, G_A77 and G_A88. The experimentally determined Φ values were divided into three categories and reported on the structure of G_A variants using the following color code: red, 0 < Φ < 0.30; magenta, 0.30 < Φ < 0.70; blue, 0.70 < Φ < 1. A conserved nucleus between the helices α 1 and α 2 is clearly evident in all the G_A variants.

G_B proteins.

The equilibrium unfolding transitions of G_B30, G_B77 and G_B88 are reported in Figure 4.11. In all cases, we observed a monotonic sigmoidal transition, suggesting the absence of equilibrium intermediates.

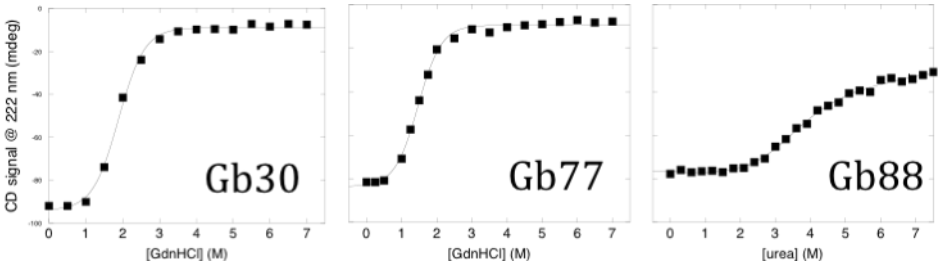


Figure 4.11. Equilibrium denaturation of G_B variants monitored by CD in 50mM sodium phosphate buffer at pH 7.2. As detailed in the text, because of the very different thermodynamic stabilities of the proteins, the equilibrium denaturations were performed at different experimental conditions: the unfolding denaturations of G_B30 and G_B77 were carried out at 25°C using GdnHCl; and G_B88 experiment was performed at 25°C using urea and 0.4M sulfate.

The folding pathway of the G_B proteins is inherently more complex than that of the G_A partners. As previously discussed in Gb1 protein, a curvature is detected only at high concentrations of GdnHCl, giving the opportunity to see experimentally both the early and late folding events. Consequently, we resorted to study the folding of G_B30, G_B77 and G_B88 at 25 °C and pH 7.2 in 50 mM sodium phosphate buffer, using GdnHCl as a denaturing agent. Unfortunately, however, in the case of G_B88, many of its site-directed variants were poorly soluble at moderate concentrations of GdnHCl and we could not obtain reliable folding data on this system with this denaturant. Furthermore, the low stability of this variant did not allow us to perform a complete Φ-value analysis in the absence of the stabilizing agent sodium sulphate. All folding studies of G_B88 were therefore carried out using urea as denaturant and in the presence of 0.4 M sodium sulphate.

The chevron plots for G_B30, G_B77 and G_B88 as a function of denaturant concentration, at 25 °C and pH 7.2, are reported in Figure 4.12. It is evident that, in analogy to what previously observed for GB1, both G_B30 and G_B77 display a pronounced curvature in their unfolding arms, as expected for a three-state folding mechanism (Parker et. al., 1995; Wildegger et. al., 1997). In the case of G_B88, while we could not detect such a curvature when performing the experiments in urea, data recorded in the presence of high concentrations of GdnHCl were similar to those of G_B30 and G_B77 (Figure 4.11). This observation suggests that all the G_B variants appear to conform to a three state folding mechanism.

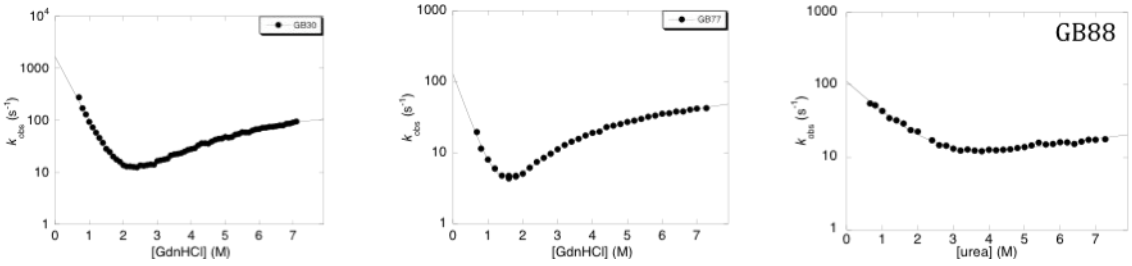


Figure 4.12. Chevron plots of G_B variants. Semilogarithmic plot of the observed rate constant for folding and unfolding of G_B30, G_B77 and G_B88 versus [denaturant] measured at pH 7.2.

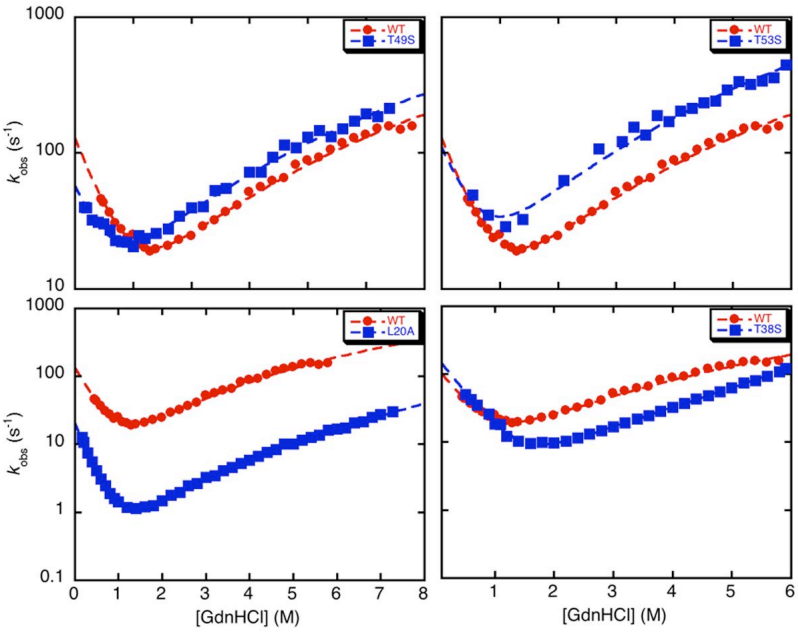


Figure 4.13. Chevron plots of wild-type (red filled circles) and representative mutants (blue filled squares) of G_B88. Observed rate constants were measured as a function of GdnHCl at pH 7.2 and 25°C. As detailed in the text, data recorded in the presence of high concentration of GdnHCl were similar to those of G_B30 and G_B77. Lines are the best fit to a three-state model.

We addressed the structural features of the early and late transition states for folding of the G_B proteins by Φ-value analysis. We produced a total of 82 mutants: 24 for G_B30, 27 for G_B77 and 31 for G_B88. 16 variants expressed poorly or were too unstable to be included in the analysis. The remaining 66 were subjected to equilibrium and kinetic folding experiments (Figures 4.14, 4.15, 4.16). In the case of G_B30 and G_B77, the data for each variant were globally fitted to a three-state equation with *m*-values assumed to be the same as those of the wild-type proteins. All data were in agreement with a three state mechanism involving the presence of two folding transition states, a more denatured-like *TS1* (β=0.77 and β=0.69 for G_B30 and G_B77, respectively) and a native-like *TS2* (β =0.97 for both the proteins). On the other hand, in the case of G_B88, because of the low solubility of its variants in the presence of GdnHCl, we could perform the experiments only in the presence of urea, obtaining structural information only on the early transition state *TS1* (β =0.81).

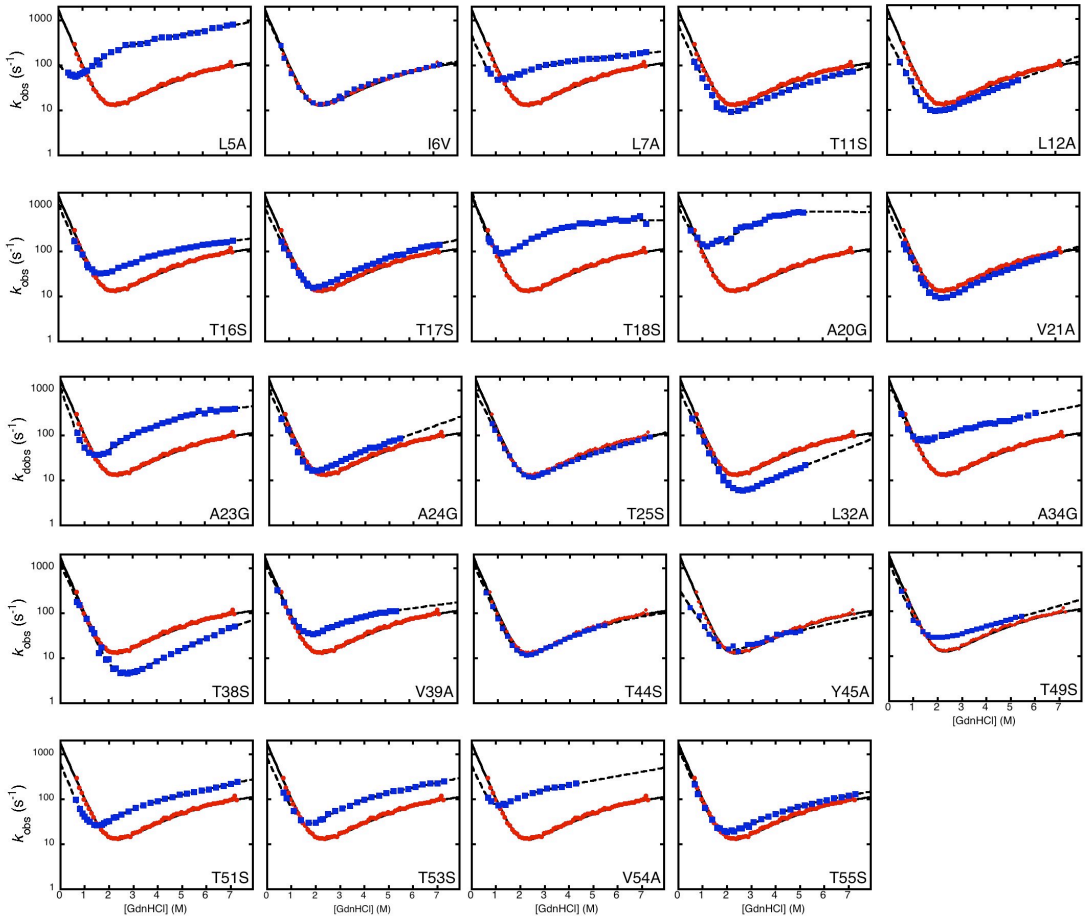


Figure 4.14. Chevron plots of wild-type (red filled circles) and mutants of G_B30. Observed rate constants were measured as a function of GdnHCl at pH 7.2 and 25°C. All the chevron plots showing a roll-over effect, hallmark of an on-pathway intermediate, were fitted to a three-state equation. Lines are the best fit to this model.

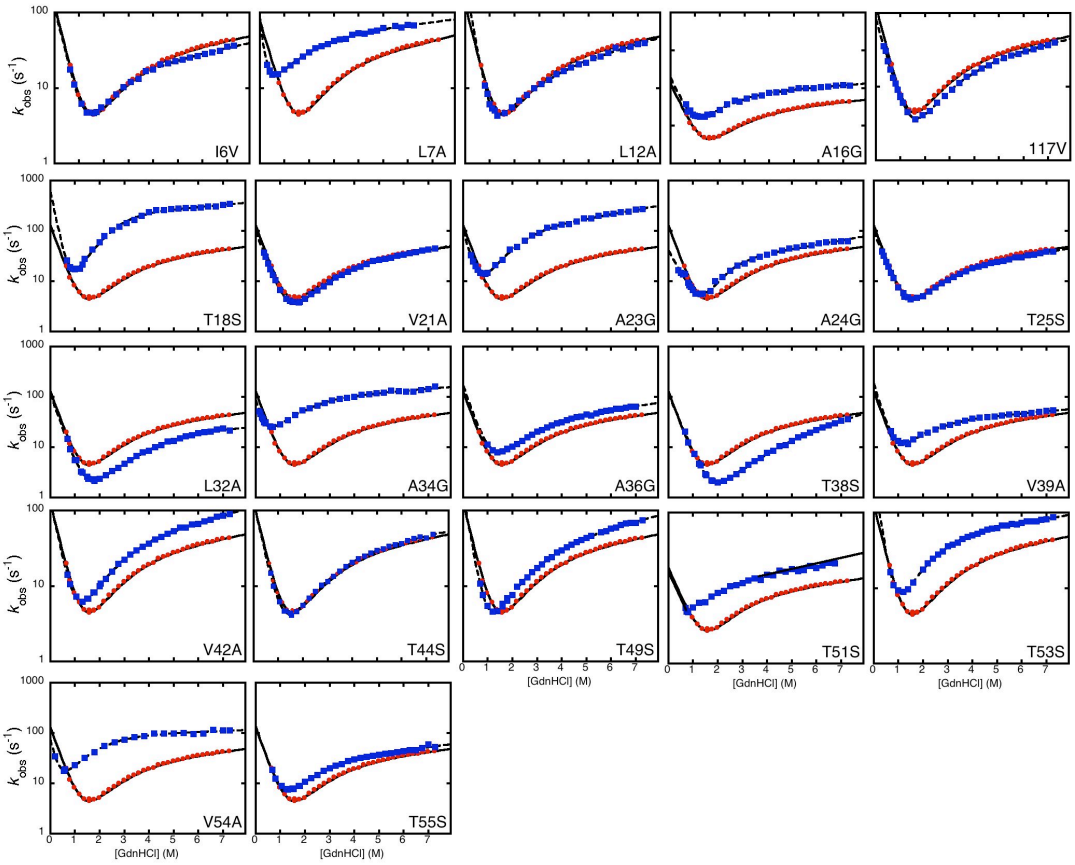


Figure 4.15. Chevron plots of wild-type (red filled circles) and mutants of G_B77. Observed rate constants were measured as a function of GdnHCl at pH 7.2 and 25°C. All the chevron plots showing a roll-over effect were fitted to a three-state equation. Lines are the best fit to this model.

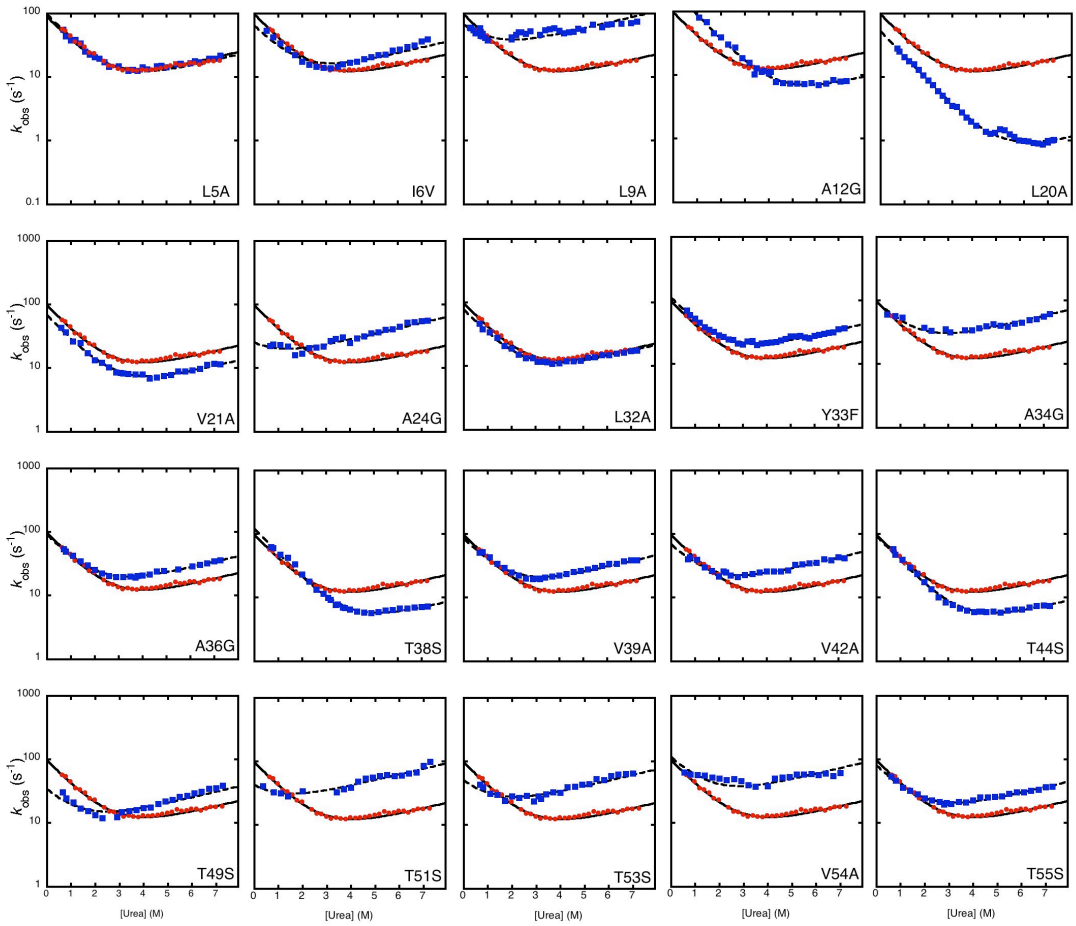


Figure 4.16. Chevron plots of wild-type (red filled circles) and mutants of G_B88. Observed rate constants were measured as a function of urea at pH 7.2 and 25°C in presence of 0.4 M sulfate. Lines are the best fit to a two-state model.

A graphic depiction of the structural distribution of the measured Φ-value for *TS1* and *TS2* of G_B30 and G_B77 and of *TS1* of G_B88 is reported in Figure 4.15. It is interesting to note that the distribution of the measured Φ-values for the first transition state (Φ_{TS1}), plotted on the native structure of G_B, show considerable differences among G_B30, G_B77 and G_B88. In fact, looking at G_B30 and G_B88, a shift of the medium-high Φ-values from the first β-hairpin to the second, with G_B77 displaying an intermediate behaviour (Figure 4.17), may be appreciated. This trend indicates that alternative folding nuclei, located at the hairpins between either β1-β2 or β3-β4, drive the folding to the G_B-topology. These nuclei may be selectively stabilized depending on amino acid

composition. Remarkably, in the native-like transition state *TS2*, which we could infer only for G_B30 and G_B77, both nuclei appear in the process of being folded and the two transition states display a similar overall structure, indicating that the alternative folding pathways converge as the native state is approached. In order to compare the Φ-values obtained under different conditions, kinetic experiments of some mutants were performed with both guanidine and urea. The calculated Φ-values were approximately the same; the folding mechanism of the G_B proteins is therefore not affected by the nature of the denaturant agent used to obtain the data (see Figure 4.18 in which the structural distribution of the measured Φ-values for *TS1* of G_B77 obtained in presence of GdnHCl (A) and urea (B) is represented).

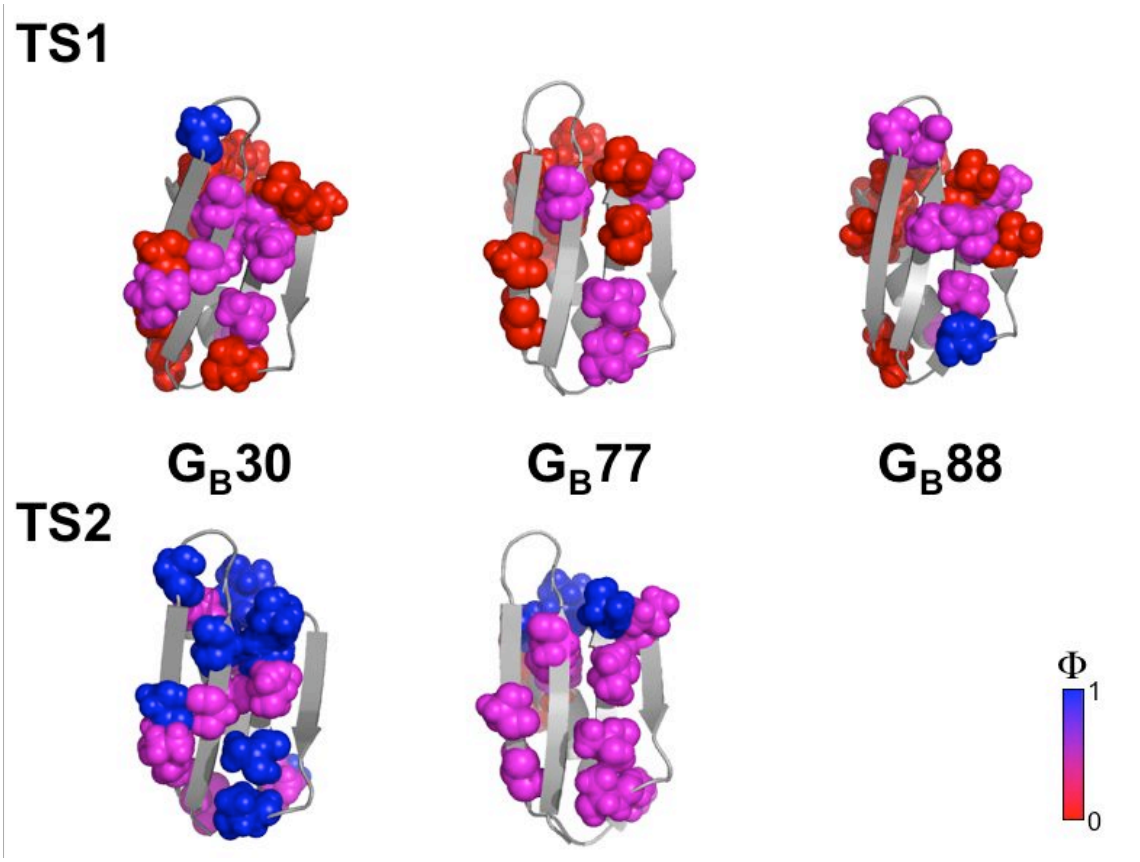


Figure 4.17. Structural distribution of the Φ-values, measured for the first transition state, on the native structures of G_B30, G_B77 and G_B88. The experimentally determined Φ values were divided into three categories and reported on the structure of G_B variants using the following color code: red, $0 < \Phi < 0.30$; magenta, $0.30 < \Phi < 0.70$; blue, $0.70 < \Phi < 1$. As detailed in the Discussion, by considering G_B30 and G_B88, it is possible to

observe, a shift of the medium-high Φ-values from the first β-hairpin to the second, with G_B77 displaying an intermediate behavior.

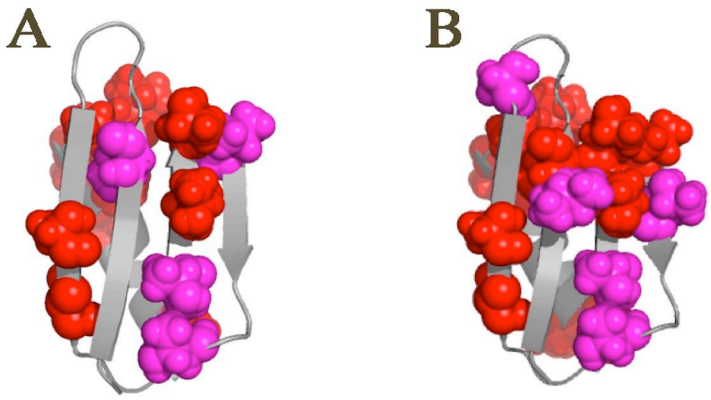


Figure 4.18. Structural distribution of the Φ-values for the first transition state of GB77 measured in presence of GdnHCl (A) and urea (B) The experimentally determined Φ-values were divided into three categories and reported on the structure of GB77 using the following color code: red, $0< \Phi <0.30$; magenta, $0.30 < \Phi< 0.70$; blue, $0.70 < \Phi < 1$.

4.2: Folding and Binding characterization of KIX and its interacting partner Trans Activation Domain of the hematopoetic transcription factor cmyb, an IDP:

4.2.1 Folding Mechanism of Kix Domain:

The dependence of the folding and unfolding rate constants of KIX versus denaturant concentration are reported in Figure 4.2.1. Data are superposed with those previously measured in reference (Horng et. al., 2005). It may be noticed that, while there is an excellent overall agreement between the data set obtained by this laboratory with those previously published, a clear difference could be observed in the region of the folding branch recorded at low denaturant concentration (i.e. at [urea] < 1.5 M). Prompted by such deviation, to circumvent the intrinsic limitations of standard stopped-flow instruments, we resorted to complement the experimental data set with experiments carried out with an ultra-rapid in-house built continuous flow instrument. A typical refolding trace of KIX measured in the presence of 10mM phosphate and 150mM NaCl buffer at pH 7.5 and 0.6 M [urea] is reported in Figure 4.2.2A.

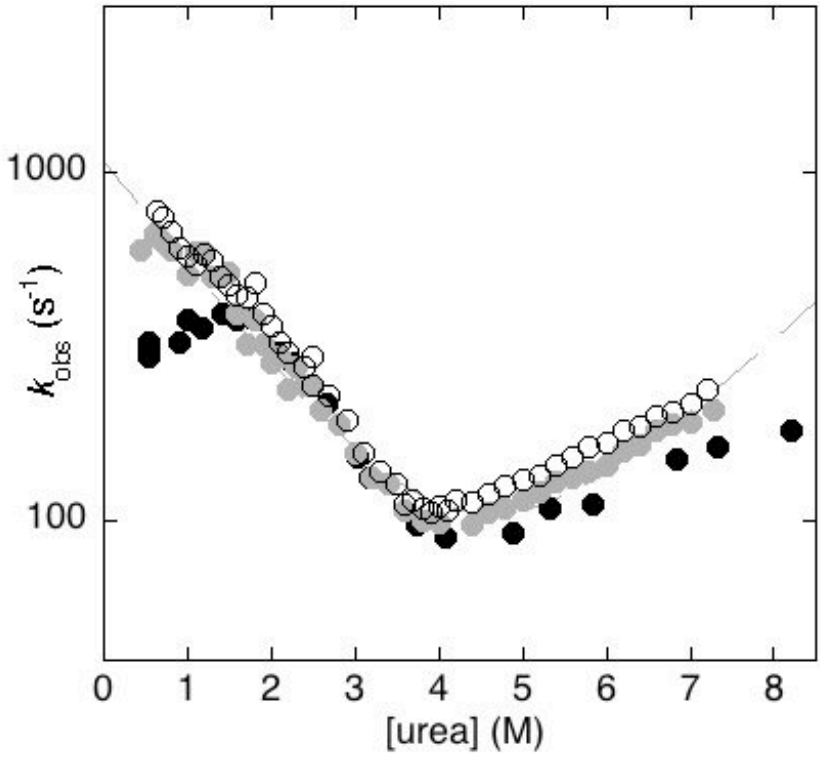


Figure 4.2.1 Chevron plot of KIX obtained by stopped-flow. Semilogarithmic plot of the observed rate constants for folding and unfolding of KIX versus [urea] at pH 7.5 at 25 °C in 10 mM NaPi and 150 mM NaCl. The chevron plots of Kix protein with His-Tag (grey circles) and without His-Tag (black open circles) obtained by stopped-flow are superposed with that previously published by Horng et al. (black filled circles).

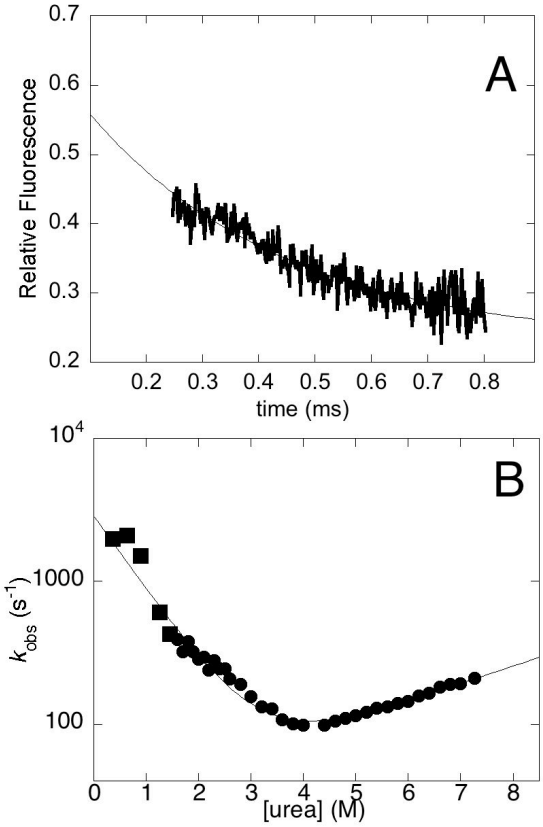


Figure 4.2.2 **Folding kinetics of KIX measured by stopped flow and continuous-flow.** A) Refolding trace of KIX measured by continuous-flow apparatus. The trace was obtained in 10mM Sodium Phosphate and 150mM NaCl buffer at pH 7.5 and in presence of 0.6 M urea concentration. B) Semilogarithmic plot of the observed rate constants for folding and unfolding of KIX versus [urea] at pH 7.5 in 10mM NaPi and150mM NaCl obtained by stopped- (●) and continuous-flow (■).

The complete chevron plot from stopped- and continuous-flow experiments is reported in Figure 4.2.2 B. It may be noticed, that, on the contrary of what previously observed, the roll-over is completely lost when the refolding rate constant is measured with an instrument characterized by a much lower dead-time. This observation suggests that previous observation of roll-over effect is most likely due to the intrinsic limitations of the stopped-flow methodology. Furthermore, an underestimation of the refolding rate constant

would correspond to an artifactual burst-phase in the amplitude analysis (an underestimated rate constant leads to an underestimated amplitude).
In an effort to further investigate the presence/absence of a roll-over effect in the folding of KIX, we carried out folding experiments in the presence of a stabilizing salt, namely in the presence of 0.4 M sodium sulfate (Figure 4.2.4). As expected, the presence of 0.4 M sulfate results in a shift of the denaturation midpoint to higher concentration (i.e. 4 M in the absence of 0.4 M sodium sulfate; 5 M in its presence). Furthermore, the observed folding rate constants are increased by a factor of 3. In analogy to what observed in the absence of sodium sulfate, when the data are recorded using both the stopped- and the continuous-flow methodology, all kinetic data displayed mono-exponential behavior with no indications of refolding roll-over effect (a typical refolding trace of KIX recorded in presence of 0.4 M sodium sulfate is reported in Figure 4.2.3). In turn, contrary to what previously reported, the chevron plot of KIX appears to satisfy a simple two-state folding without the accumulation of intermediates.

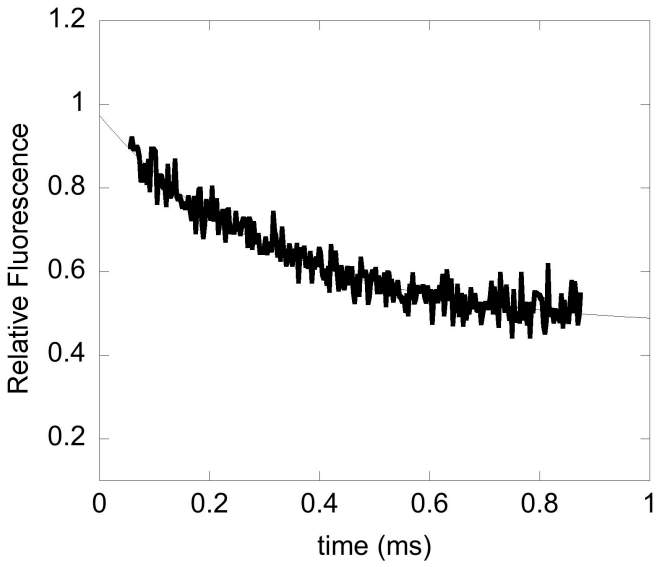


Figure 4.2.3 **Refolding trace of KIX recorded by continuous-flow**. The trace was obtained in 10mM Sodium Phosphate, 150mM NaCl, and 0.4 M Na_2SO_4 buffer at pH 7.5 in presence of 0.54 M urea concentration.

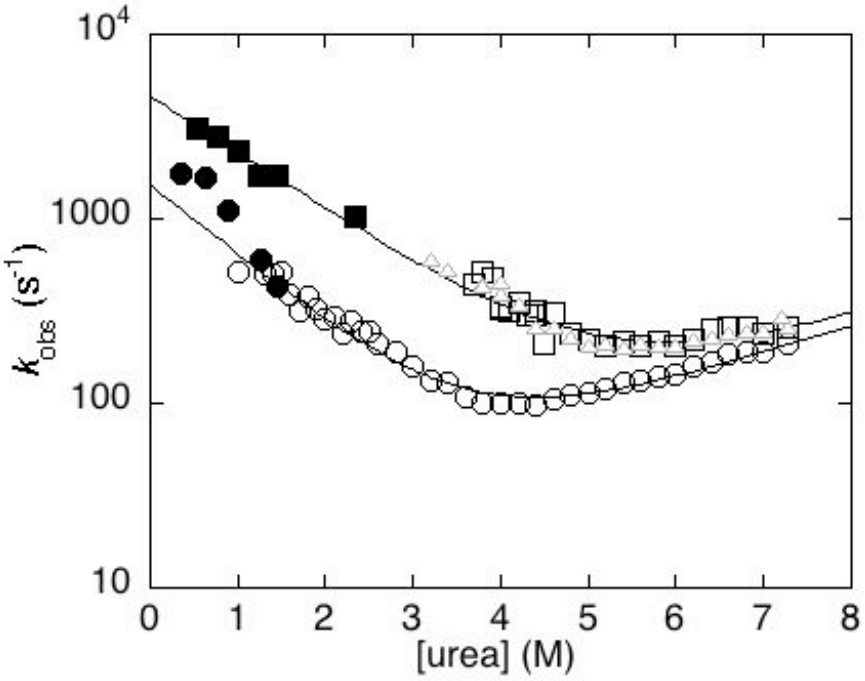


Figure 4.2.4 **Folding kinetics of KIX measured in presence of sodium sulphate.** Semilogarithmic plot of the observed rate constants for folding and unfolding of KIX versus [urea] at pH 7.5 in 10mM NaPi and 150mM NaCl obtained in presence of 0.4 M Na_2SO_4 by stopped-(❍) and continuous-flow (●) and in its absence by stopped-flow (■). Stopped flow kinetics of KIX protein without hexa histidine tag is also reported here (Δ).

4.2.2 Folding and Binding characterization of KIX and cMybTAD:

To address the structural properties of the transactivation domain of c-Myb, we carried out equilibrium circular dichroism (CD) experiments. The far-UV CD spectrum of c-Myb, reported in Fig. 4.3.1, shows absence of secondary structure, confirming that c-Myb is an IDP (Zor et. al., 2004).
The molecule 2,2,2 trifluoroethanol (TFE), a well-known helix stabilizer, is routinely used to induce folding in peptides lacking structure in water (Jasanoff et. al., 1994; Myers et. al. 1998). Because c-Myb is helical once bound to its physiological partner KIX, we resorted to use TFE to induce its folding in the free state. The CD spectra of c-Myb recorded in the presence of different concentrations of TFE, reported in Fig. 4.3.1b, show that c-Myb undergoes a folding transition with an apparent midpoint at 18% v/v TFE. The dependence of the CD signal upon [TFE] conforms to a standard two-state transition and allows calculating a folding free energy in water, $\Delta G = 2.5 \pm 0.5$ kcal mol^{-1}.

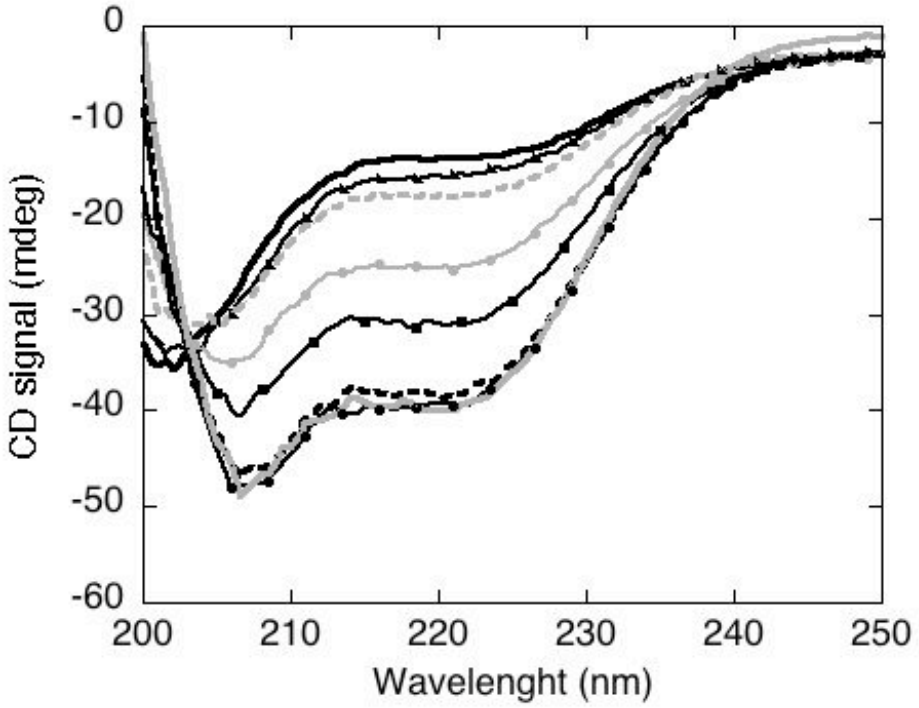


Figure 4.3.1.
A) CD spectra of c-Myb measured in 50mM sodium phosphate buffer 150mM KCl at pH 7.2 using a 1-cm cylindrical cuvette at 10°C with [protein] = 10 μM in presence of different concentration of TFE. Spectra are depicted in a grey scale varying from black (in the absence of TFE) to light grey (in the presence of 40% TFE).

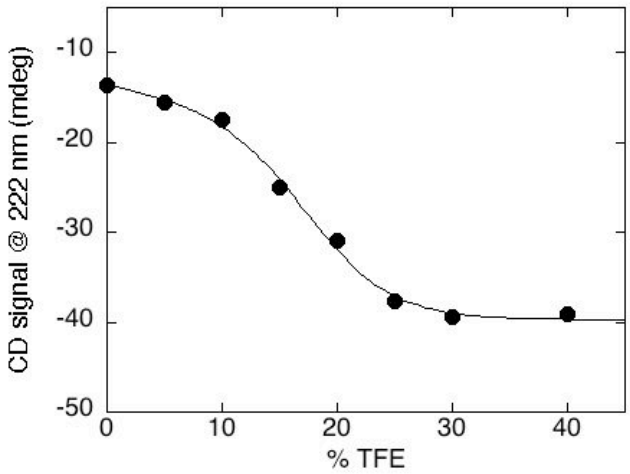


Figure 4.3.1B) CD signal of cmyb upon [TFE] monitored at 222 nm in 50mM sodium phosphate buffer, 150mM KCl at pH 7.2 and 10°C. Line is the best fit to a standard two state transition.

The mechanism of binding of c-Myb to KIX.

To study the mechanism of recognition between KIX and c-Myb, we produced different constructs characterized by a detectable fluorescence change upon binding. We first introduced a tryptophan residue, i.e. a good fluorescent probe, in the proximity of the binding pocket and produced a pseudo-wild-type Y72W variant, called pwtKIX (residue 72 is highlighted in Figure 4.3.2). The equilibrium binding transition of c-Myb to pwtKIX, monitored by fluorescence, is consistent with a simple hyperbolic behaviour, with an apparent K_D of 20 ± 4 μMm measured at pH 7.2 in the presence of 150 mM KCl and 25 °C (Figure 4.3.3).

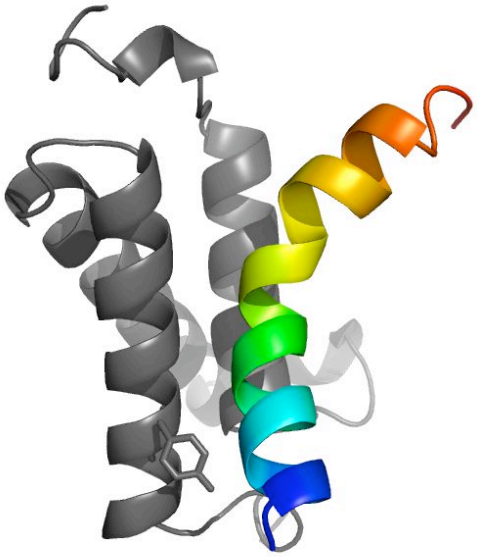


Figure 4.3.2. Three dimensional structure of the complex between KIX and c-Myb. Residue Tyr72, mutated to Trp in pwtKIX, is highlighted in sticks.

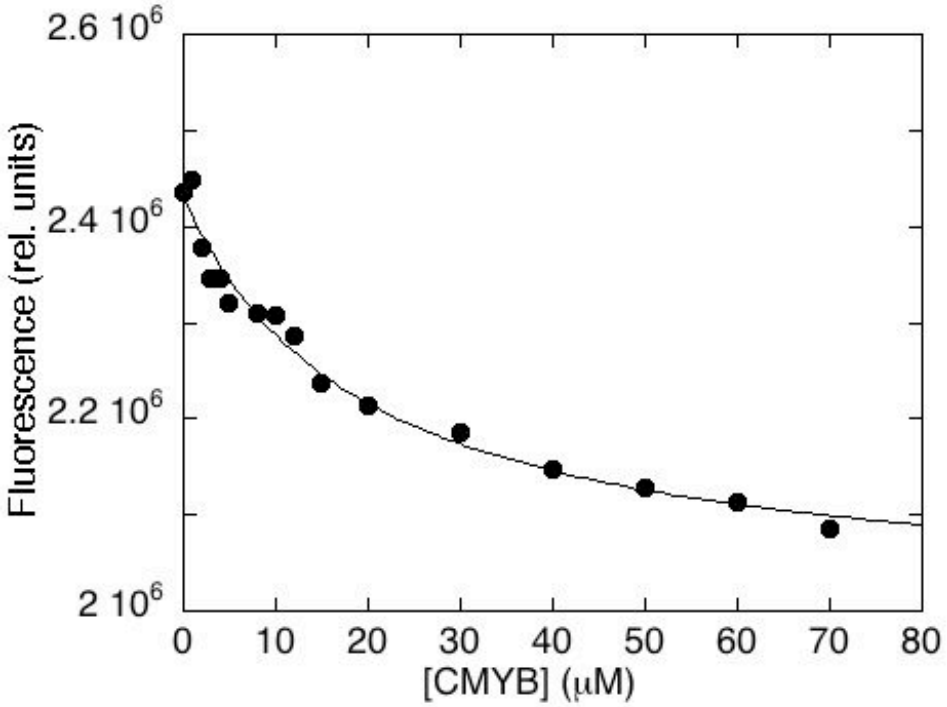


Figure 4.3.3 Equilibrium binding transition of c-Myb to pwtKIX monitored by fluorescence in 50mM sodium phosphate buffer,150mM KCl at pH 7.2. The emission wavelenght was 370 nm (excitation 280 nm). pwtKIX concentration was 1μM. Line is the best fit to a simple hyperbolic behavior.

We studied the kinetics of binding of c-Myb to pwtKIX using a temperature jump discharge capacitor apparatus (TGK Instruments, UK). A mixture containing a constant concentration of pwtKIX (10 μM), incubated with different concentrations of c-Myb, was subjected to a 9 °C rapid temperature jump, with a shift from 16 to 25 °C. Under all conditions, the observed relaxation kinetics were consistent with a single exponential time-course, suggesting the absence of detectable intermediates. The observed dependence of the relaxation rate constant on c-Myb concentration is reported in Fig. 4.3.4. Despite of the inherent complexity of the folding-and-binding recognition reaction between pwtKIX and c-Myb, the observed relaxation behavior displays a simple second-order linear dependence. Furthermore, the apparent K_D of 30 ± 3 μM that can be calculated from the kinetic experiments

is consistent with the value obtained by equilibrium, supporting the lack of detectable intermediates.

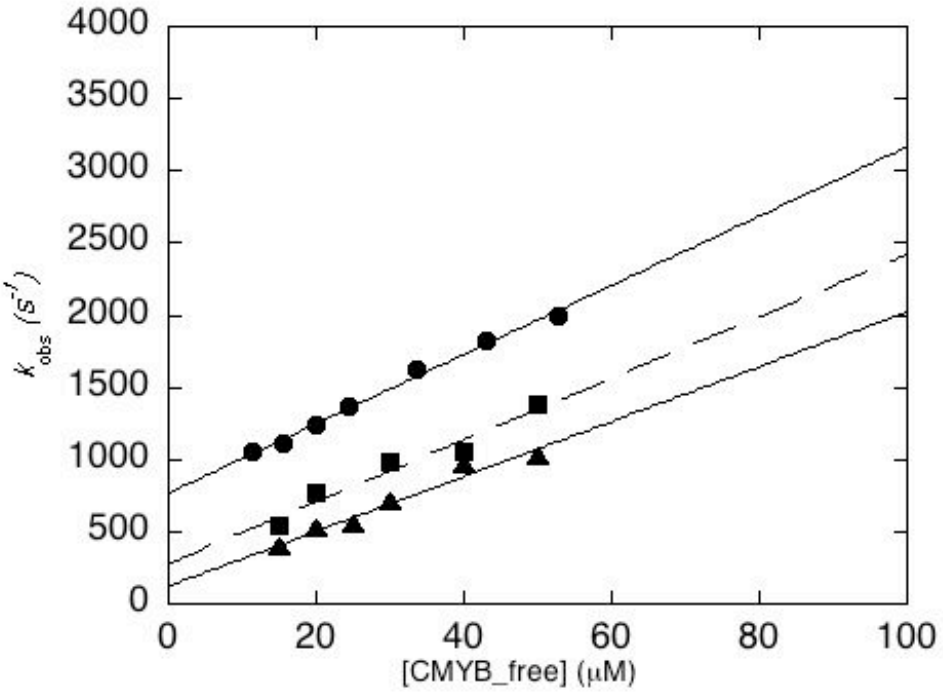
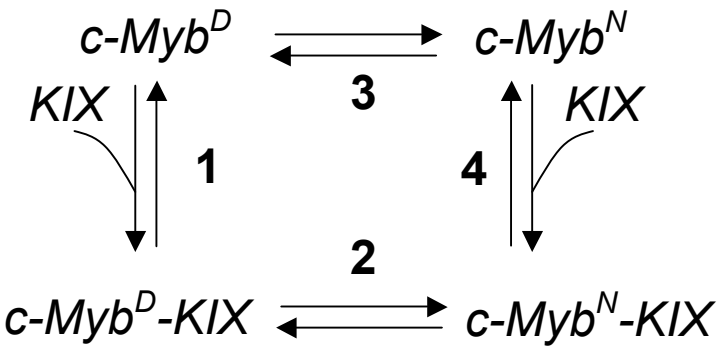


Figure 4.3.4 Observed dependence of the relaxation rate constants measured by T-jump apparatus in 50mM sodium phosphate buffer,150mM KCl at pH 7.2, at different [c-Myb] in absence (●) and in presence of 5%TFE (■) and 7% TFE (▲).Temperature was rapidly changed from 16 °C to 25°C with a jump-size of 9°C. The observed rate constants display a pseudo-first order linear dependence.
B) Dependence of the activation energy for association and dissociation of the c-Myb-pwtKIX complex on the [TFE] (■) compared to the effect of the equilibrium free energy of folding of c-Myb, as calculated from the equilibrium TFE tritation reported in Fig.4.3.1 (●).

To shed light on the function, is any, of disorder in IDP systems would require to clarify the order of events between folding and binding: does folding occur prior binding or do IDPs recognize their partner in a disordered conformation and fold only after binding? A monomeric protein undergoing a ligand induced conformational change can be described by a square mechanism:



(Scheme 1)
Where c-Myb^D and c-Myb^N denote the denatured and folded conformations of c-Myb, respectively. Complex formation progressing through pathway 1 and 2 is representative of an induced-fit model, whereby ligand binding induces a conformational change (Koshland et. al., 1966), i.e. folding-after-binding. Alternatively, binding progressing through pathway 3 and 4 assumes that two alternative conformations of c-Myb are in pre-equilibrium in the absence of the ligand, formally similar to a concerted model (Monod et. al., 1965), or folding-before-binding.
To distinguish between the two different scenarios, we performed T-jump experiments in the presence of increasing concentrations of TFE, which stabilizes the helical conformation of c-Myb (Figure 4.3.1). Remarkably, inspection of Figure 4.3.5 reveals that, while the presence of TFE has a negligible effect on the apparent association rate constant, it stabilizes the c-Myb-KIX complex by lowering the dissociation rate constant. In these experiments, the highest concentration of TFE employed was 7% (v/v), because we found pwtKIX to precipitate at higher concentrations. In analogy to the so called Φ-value analysis in protein folding studies (Fersht et. al.,

1992), the lack of stabilization of the rate limiting step suggests the folding transition state to be as unstructured as the denatured c-Myb. Consequently, it may be concluded that folding of c-Myb is a late event occurring rapidly after the main rate-limiting barrier and yielding the observed fluorescence quenching (Fig. 4.3.3), the overall reaction being consistent with a folding-after-binding mechanism.

A powerful method to infer quantitatively the effect of a perturbation on the reaction mechanism is to compare the effect on activation free energy with that on equilibrium (Leffler, 1953). Figure 4.3.5 depicts the dependence of the activation energy for the dissociation of the c-Myb-KIX complex on the [TFE], compared to the effect of the equilibrium free energy of folding of c-Myb, as calculated from the TFE titration reported in Figure 4.3.2. The linear dependence of the two parameters on [TFE] is essentially the same, while the association rate constant is nearly insensitive to [TFE]. This finding strongly suggests that the stabilization of c-Myb in its helical folded conformation may account for all the stability the c-Myb-KIX complex. Furthermore, because the stabilization is due to a decrease of the dissociation rate constant, it may be concluded that the folding of c-Myb is a late step that occurs only downhill the main limiting barrier, i.e. c-Myb recognizes KIX in a disordered non-helical conformation.

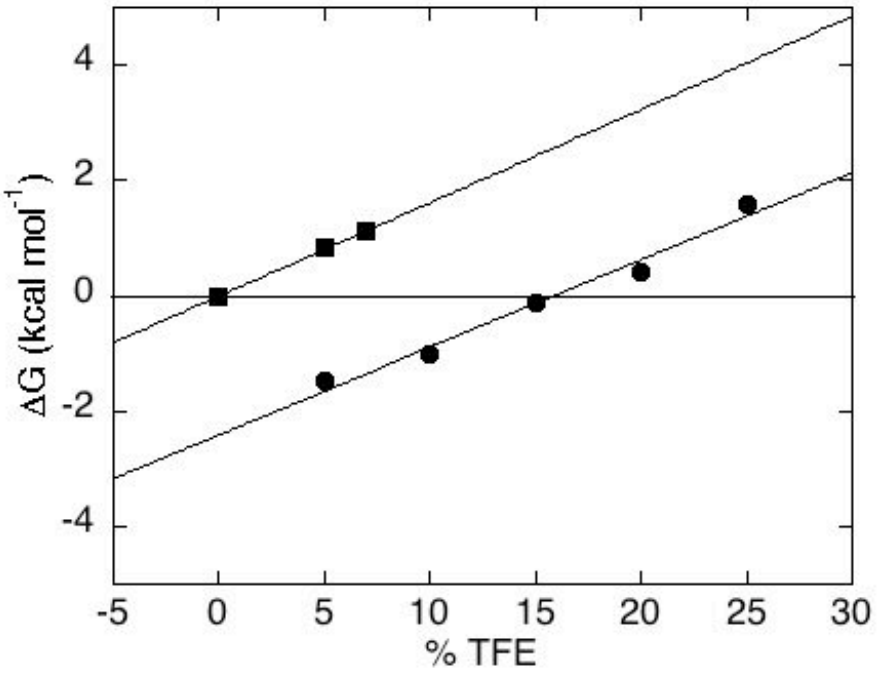


Figure 4.3.5 Dependence of the c-Myb free energy calculated for the logarithm of the apparent association (●) and dissociation (■) rate constants measured at different pH conditions. The apparent dissociation rate constants were obtained by a displacement experiment consisting in a rapid mixing of a pre-incubated complex between pwtKIX-c-Myb* with an excess of wtKIX. Lines are the best fit to an equation implying a single protonation site.

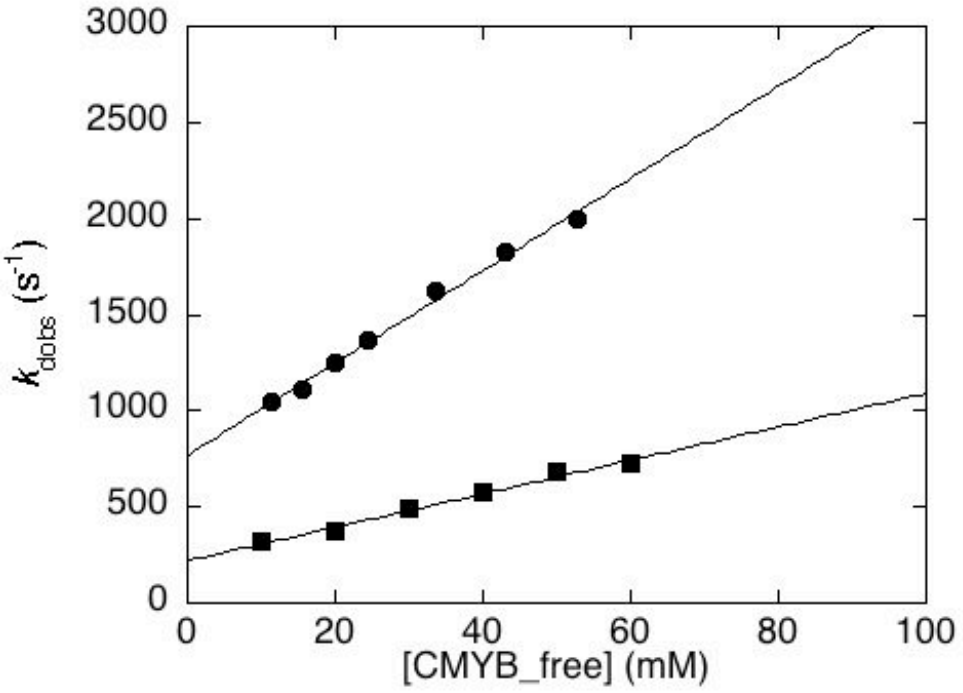


Fig. 4.3.6. Observed rate constants for the binding of pwtKIX to c-Myb (●) and the engineered c-Myb* (■) construct containing a pro-domain of subtilisin at its N-terminal. The experiments were carried out in 50mM sodium phosphate buffer, 150mM KCl at pH 7.2 and 10°C. pwtKIX concentration was 1μM.

Effect of pH on the binding kinetics

In order to produce c-Myb in high yields using *E.coli* expression strains, we engineered a construct where the transactivation domain of c-Myb was fused to the pro-domain of subtilisin, denoted as c-Myb*. While this construct binds to pwtKIX with an overall affinity comparable to that of free c-Myb, an investigation of its binding kinetics (Figure 4.3.6) revealed a clear-cut effect on both the association and dissociation rate constants. These relatively

minor changes are most likely due to some steric hindrance induced by the presence of the large tag, but do not affect the K_D for the binding to KIX. Furthermore, c-Myb* displayed the additional advantage of a much better signal to noise ratio compared to free c-Myb. Because of such improved signal to noise due to the tag, we resorted to expand the analysis of the mechanism of binding using c-Myb* and explored the effect of pH from 4 to 9.

The dependence of the observed rate constant on the concentration of c-Myb*, depicted in Figure 4.3.7, shows that pH affects both the intercept and the slope, the overall behaviour nevertheless being consistent with a simple bimolecular reaction. To validate the pH dependence of the apparent dissociation rate constant, we carried out a full set of displacement kinetic experiments where a pre-incubated complex between pwtKIX and c-Myb* was rapidly mixed with an excess of wild-type KIX (i.e. the wild type domain with a Tyr at position 72). Under appropriate conditions, the reaction is rate limited by the dissociation of the pre-existing complex, with the excess of wild-type KIX substituting for the bound pwtKIX in the complex with c-Myb*. The dissociation process was measured at different concentrations of wild-type KIX, ranging from 2 to 10 fold, and the observed rate constants were found to be insensitive to its concentration at all pH values (Figure 4.3.8). This methodology allows estimating unequivocally the dissociation rate constants, in analogy to classical experiments on myoglobin (Antonini and Brunori, 1971).

A plot of the dependence of the apparent association and dissociation rate constants on pH is reported in Figure 4.3.9. Both rate constants increase with pH in a sigmoidal profile, consistent with the protonation of a single group; the apparent pKa's for the association and dissociation rate constants are 4.2 and 7.6, respectively.

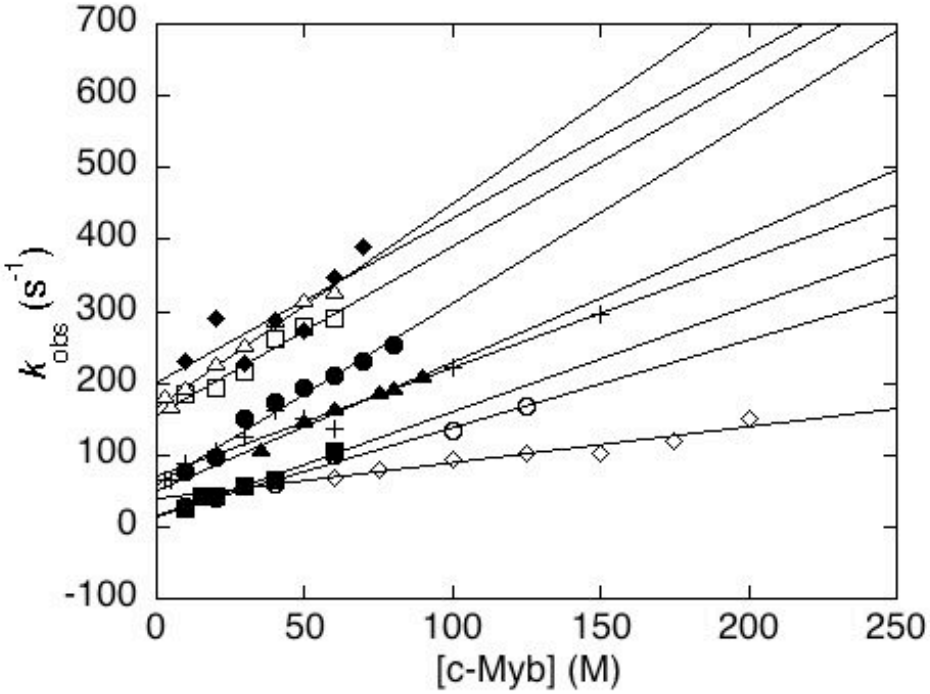


Fig. 4.3.7. Observed rate constants for the binding of pwtKIX to c-Myb* measured under pseudo-first order conditions in presence of 150mM KCl at different pH conditions (◊ pH 4.0, ○ pH4.5, ■ pH 5.0, ▲ pH 5.5, + pH 6.0, ● pH 7.2, □ pH 8.0, △ pH8.5, ◆ pH 9.0). pwtKIX concentration was 1μM.

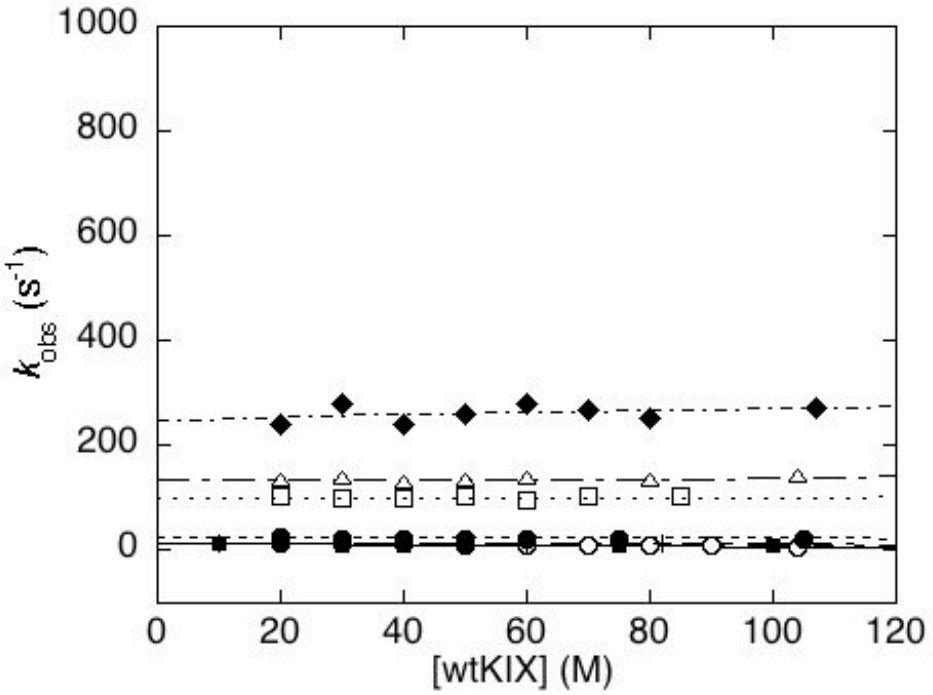


Fig. 4.3.8. Observed rate constants for the dissociation of the complex between pwtKIX and c-Myb* measured by rapidly mixing the pre-incubated complex with different concentrations of wtKIX under pseudo-first order conditions. The experiments were carried out in presence of 150mM KCl and 1mM DTT at different pH conditions (○ pH4.5, ■ pH 5.0, + pH 6.2, ● pH 7.2, □ pH 8.0, △ pH8.5, ◆ pH 9.0).

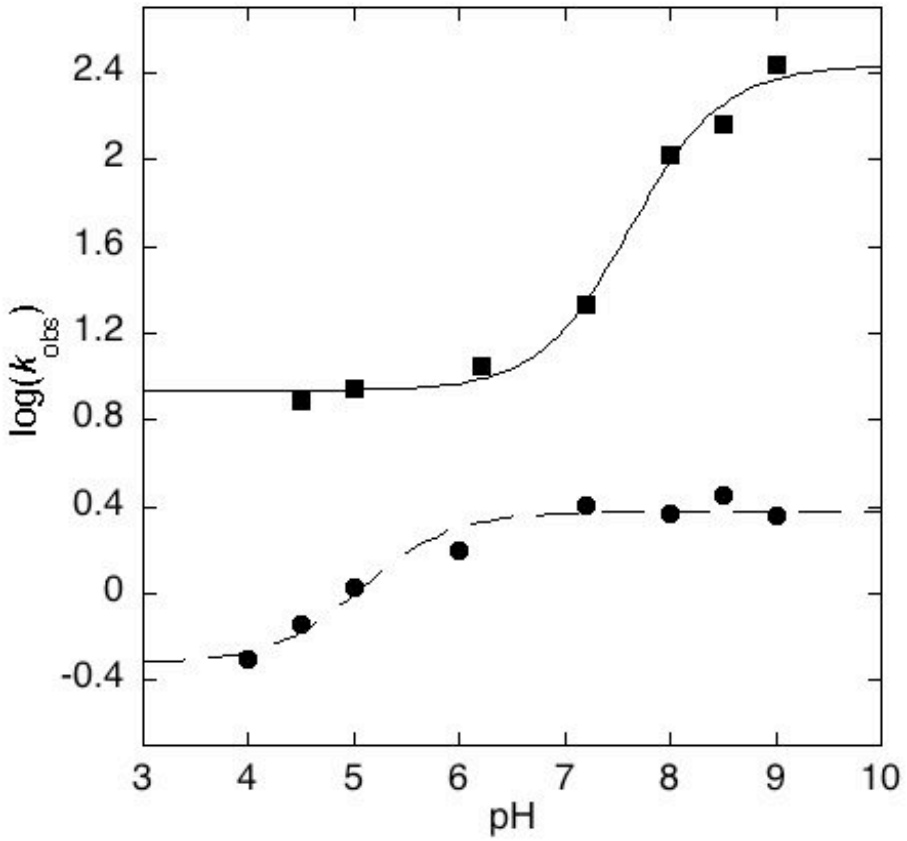


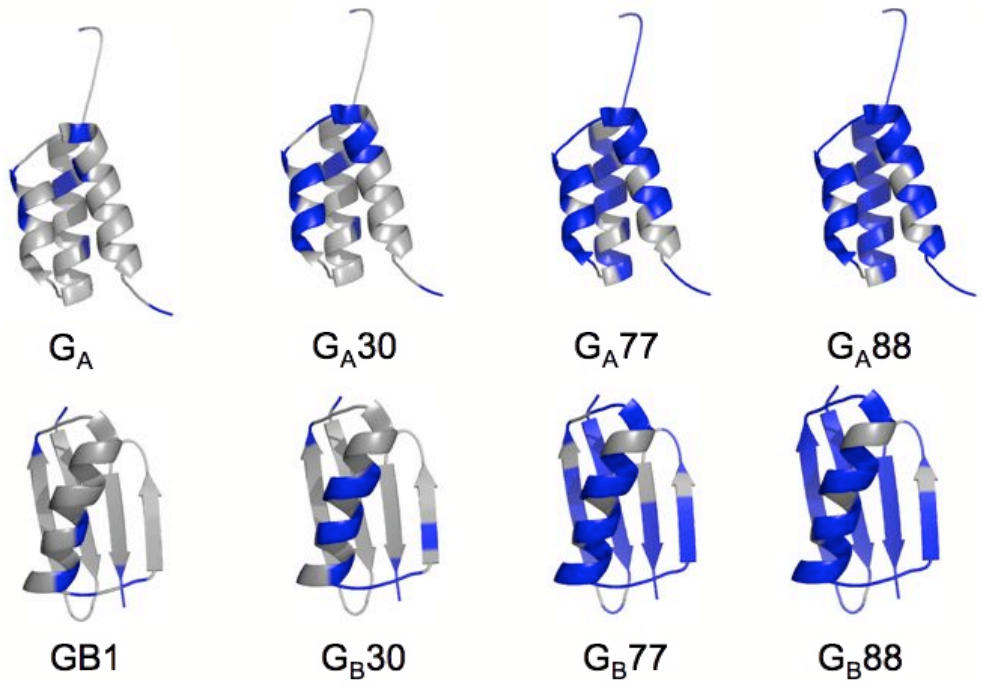
Figure 4.3.9. Dependence of the logarithm of the apparent association (●) and dissociation (■) rate constants measured at different pH conditions. The apparent dissociation rate constants were obtained by a displacement experiment consisting in a rapid mixing of a pre-incubated complex between pwtKIX-c-Myb* with an excess of wtKIX. Lines are the best fit to an equation implying a single protonation site.

DISCUSSION

Chapter 5: Discussion

5.1 Nearly identical yet different; the folding pathway of G_A and G_B proteins:

To see the molecular determinants of heteromorphic pairs of proteins with increasing
degree of sequence identity we carried out a systematic and very extensive analysis of each of the heteromorphic variants designed by Philip N. Bryan (Alexander, et. al., 2007; He, et. al., 2008), including the natural G_A and G_B domains. The natural G_A domain does not contain any intrinsic fluorescence probes (i.e. no Trp) and differs by only three amino acids from the variant G_A30. Therefore we omitted its characterization from our analysis and decided to focus on the folding pathway of GB1, a popular system for protein folding studies, as well as the set of heteromorphic pairs G_A30 and G_B30, G_A77 and G_B77, G_A88 and G_B88 (Figure 5.1).
The extensive experimental results obtained for the different domains are briefly discussed below.




```
GA    MEAVDANSLA QAKKAAIKEL KQYGIGDYTI KLINNAKTVE GVESLKNEIL KALPTE
GB1   MTYKLILNGK TLKGETTTEA VDAATAEKYF KQYANDNGVD GEWTYDDATK TFTVTE

GA30 MEAVDANSLA QAKEAAIKEL KQYGIGEKYI KLINNAKTVE GVWSLKNEIL KALPTE
GB30 MTYKLILNGK TLKGETTTEA VDAATAEKYF KLYANDKTVE GEWTYDDATK TFTVTE

GA77 TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL KATVTE
GB77 TTYKLILNGK QLKEEAITEA VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE

GA88 TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL TFTVTE
GB88 TTYKLILNLK QAKEEAIKEL VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE
```

Figure 5.1. Structures and sequence alignments of the different G_A and G_B variants. All engineered G_A and G_B proteins, designed by Bryan and co-workers, display structure and function similar to their respective natural wild-type domains G_A and GB1. For each protein, amino acid identities are shown in blue and non-identities in gray.

5.1.1 Detecting an unexpected folding intermediate in the Gb1 protein:

The folding pathway of the small α/β protein GB1 has been extensively

studied during the past two decades using both theoretical and experimental approaches (Alexander, et. al., 1992; Clarke, et. al., 1999b; Krantz, et. al., 2002; Ding, et. al., 2004; Chung, et. al., 2010). Most of these studies provided a consensus view that the protein folds in a two-state manner.

In the course of our characterization of the folding of GB1, we noted that this protein, when challenged over a wide range of denaturant concentrations, appeared to display a previously undetected complex behavior (namely a 'roll-over' effect, highlighted in Figure 4.3). Because this roll-over effect was inconsistent with a simple two-state behavior (as discussed in the Introduction), we decided to study the folding pathway of GB1 under a variety of different experimental conditions, i.e. carrying out equilibrium and kinetic experiments over a wide range of pH values (from 2.0 to 9.6).
The existence of an intermediate in the folding of GB1 was previously proposed by Roder and coworkers (Park, et. al., 1997; Park, et. al., 1999), who suggested the presence of a collapsed state, accumulating in the *ms* time-range, as observed by continuous-flow ultra rapid mixing experiments. This low-energy intermediate was later questioned by Sosnick and co-workers (Krantz, et. al., 2002). Importantly, however, the partially folded state identified in this work, is distinct from that previously suggested, being a high energy species that never accumulates and whose presence suggests that GB1 folds *via* a complex and rough energy barrier with at least two discrete major transition states.
When folding is characterized by a complex chevron plot, the deviation from linearity observed may have different origins: i) the curvature may be due to movement of the position of the transition state along a single broad barrier (Oliveberg 1998; Oliveberg, et. al., 1998) or ii) the curvature may reflect a change in the rate-limiting step suggestive of a multi-state process (Bachmann and Kiefhaber 2001; Sauder J.M. 1996; Walkenhorst, et. al., 1997). In the case of GB1, the curved chevron plots are observed only above pH 6 (Figure 4.2), making the broad transition state model less likely since it would imply that it is possible i) to distort drastically the folding free-energy profile and ii) to switch between a narrow energy maximum (linear chevron plot) to a broad energy maximum (curved chevron plot) by changing pH. Furthermore, in analogy to what has been observed previously for other proteins (Bachmann and Kiefhaber 2001; Gianni, et. al., 2009; Gianni, et. al., 2005), detection of a roll-over effect only under some solvent conditions seems more consistent with a multi-step folding pathway. The observed chevron plots of GB1, obtained at different pH conditions, were fitted globally. The excellent statistical parameters of the global fit suggest that the two transition states are relatively robust and maintain their overall structural

features when solvent conditions are varied. This observation is in stark contrast with previous experimental work, which suggested the unfolding *m*-value to depend strongly on experimental conditions (McCallister, et. al., 2000; Park, et. al., 1997; Park, et. al., 1999). We conclude that both transition states *TS1* and *TS2* display a robust structure that is by-and-large maintained when solvent and/or sequence composition is altered.
It is of interest to analyze the dependence of the unfolding rate constants measured for the two transition states at different pH values. In fact, both energy barriers appear to display a monotonic transition at acidic and alkaline pH values. Both profiles are consistent with a model involving the protonation of two different groups with apparent pK_a of ~ 4 and ~ 8 (Figure 4.4). While the alkaline transition displays approximately the same change in activation free energy for both *TS1* and *TS2*, the acidic transition is more pronounced in *TS1*, suggesting the presence of a charged interaction that is weak or not formed in *TS1* and is consolidated in *TS2*. Inspection of the three-dimensional structure of GB1 suggests that such an interaction may be either Lys4-Glu15 (located at the N-terminal β-hairpin) or Lys10-Glu56 (between the N-terminal turn and the C-terminus of the protein). These considerations led to the idea that the structures of the transition states appear robust to changes in pH and may be characterized by an extended nucleus, which is stabilized by both the N- and C-terminal beta-hairpins, as well as by contacts between the N- and C-terminal strands.

5.2 Comparing the folding pathway of the G_A and G_B variants at nearly atomic resolution: Φ-value analyses.

The real goal in protein folding studies is to unveil the correlation between sequence information and reaction mechanisms. A classical approach to address this question has been to study proteins that differ in amino acid sequence but share the same fold (Calosci et. al., 2008; Chi et. al., 2007; Clarke et. al., 1999; Friel et. al., 2003; Travaglini-Allocatelli et. al., 2004). On the other hand the design and production of proteins sharing high sequence identity, yet displaying a different structure and function, allows for the first time to approach the folding problem from a complementary perspective (Alexander et. al., 2007; Alexander et. al., 2009; Parker et. al., 1995). In this work, we have characterized the complete folding pathway of heteromorphic proteins originating from two domains of streptococcal protein G being either fully α-helical (called G_A) or largely β (called G_B). An extensive Φ-value analysis of G_A protein variants (G_A30, G_A77, and GA88)

and of G_B protein variants (G_B30, G_B77 and G_B88) generates a "matrix" of proteins, which share the same topology (the two rows of Figure 5.1) and display increasingly high sequence identities while keeping a different structure.

When does a protein commit to its native topology in its folding pathway? In the present work, by systematically describing the structural features of the folding transition states of G_A30, G_A77, G_A88, as well as of G_B30, G_B77 and G_B88, we have taken our comparative study on heteromorphic pairs to the next level of complexity addressing by Φ-value analysis the structure of the transition states. Remarkably, the structural distribution of the measured Φ-values in all these variants is reminiscent of what has been previously observed in many single domain proteins, with a weak nucleus displaying moderately high values of Φ (> 0.7), and simultaneous formation of native-like extended structure, which gradually tapers off from the nucleus to other regions of the protein (Abkevich et. al., 1994; Itzhaki et. al., 1995). Accordingly, the transition states of all the proteins considered here appear to resemble a distorted version of the corresponding native states, with some polarization of structure at the N- or C-terminal β-hairpins in the case of G_B30 and G_B88 respectively. Indeed, despite the very high level of sequence identities, when comparing the folding of each heteromorphic pairs, no common intermediate was detected suggesting that they fold *via* completely independent paths. Overall, all these findings converge in indicating that proteins commit very early to their topology and the structural features leading to their native state are, most likely, already imprinted in their denatured states.

A crude but reliable method to compare the folding pathways of these different proteins is to analyze the structural distribution of measured Φ-values. From this perspective, data indicate that, whilst in the case of G_A the mechanism appears rather robust to divergence in primary structure, in the case of G_B, folding is more malleable. In fact, when mapped onto the corresponding native structures (Figures 4.10), the Φ-values reveal a conserved transition state among all the three G_A variants (G_A30, G_A77, G_A88); on the other hand, G_B30 and G_B88 clearly display a shift of the medium-high Φ-values from the first β-hairpin to the second, with G_B77 displaying an intermediate outlook (Figure 4.15). This finding may be seen in the light of the proposal that the number of accessible pathways for folding is determined by the different nucleation motifs contained within a given native topology (Lindberg and Oliveberg 2007). For example, the structure of ribosomal protein S6 seems to be composed of two different nucleation patterns acting as independent cooperative units, each of which constitutes a

separate entry to parallel folding trajectories (Haglund et. al., 2008). Accordingly, it may be suggested that in the case of G_B, the symmetrical organization of its three-dimensional topology implies the presence of multiple nucleation motifs that permit alternative folding pathways.

For our purposes, it would be revealing to understand which structural determinants preclude the sequence of the G_A proteins to adopt the structure of G_B and vice versa. Recent molecular dynamics simulations on G_A88 and G_B88 suggested the long and stable helix in the central region of the sequence of G_B88 preventing the polypeptide chain to form the loop connecting helix 1 and helix 2 in G_A88 and thus to fold into a fully helical structure (Morrone et. al., 2011). This finding is further corroborated by the high helical propensity of the only α-helix of all the G_B proteins as predicted by AGADIR (Munoz and Serrano, 1997), when compared to the G_A counterparts. By following this view, it is interesting to note that structure selection in the G_A proteins seems to be initiated by the formation of the critical contacts between helix 1 and helix 2 (Figure 4.10). In the case of the G_B proteins, on the other hand, the Φ-value analysis reported in this work, together with the previously published molecular dynamics simulations, suggests a scenario whereby folding is guided by the alternative docking of the N- or C-terminal hairpins on the central long helix, that might be (partially) pre-formed in the denatured state.

Although the protein folding reaction involves the formation and breakage of a myriad of contacts, a typical feature of small single domain proteins is the ability to fold co-operatively (Bryngelson et. al., 1995; Fersht 1995, Munoz et. al., 1997). Many weak non-covalent bonds form simultaneously and, very often, only the fully native and fully denatured states may be experimentally detected. Yet, because co-operativity is never extreme, not all residues are equally important for folding and one or more regions of the protein tend to display a selective propensity to fold independently (Fersht 1995, Munoz et. al., 1997). These sub-domains initiate the folding reaction and act as folding nuclei. Accordingly, when co-operativity decreases, the energetic coupling for folding of different regions of the same protein is also decreased, i.e. different regions of the protein may fold independently and folding may become modular. In this perspective, it is of interest to compare the different robustness of the folding mechanisms for the G_A and G_B proteins. In fact, while G_A, a two-state system, appears to display a unique nucleus, in the case of G_B, where the tendency to populate folding intermediates is documented, alternative nuclei are present and folding is more sensitive to changes in sequence composition. Overall, our data reveal that pathway malleability is determined by the presence of multiple nuclei; the segregation of such nuclei results in stabilization of folding intermediates, whereas ‘perfect’ two-state

systems are characterized by a unique diffused nucleus and, therefore, by a robust folding pathway.

5.3 Folding and Binding characterization of KIX and its interacting partner c-Myb, an IDP:

5.3.1 KIX is a two state folder:

The KIX domain of CBP has been reported to fold through an early intermediate based on two classical observations: significant burst-phase amplitudes and rollover in the folding limb of the chevron plot (Horng et. al., 2005). In carrying out a detailed chevron plot analysis by the combination of stopped- and continuous-flow experiments, both in the presence and absence of the stabilizing agent sodium sulphate (0.4M), we did not detect the previously observed signatures for such a species. Even if there are several examples in the literature of burst-phase intermediates in fast folding protein domains (Qi et. al., 1998; Shastry et. al., 1998; Ferguson et. al., 1999; Jemth et. al., 2004), this effect must be interpreted with caution. Roll-over effect and burst phase, in fact, may have alternative explanations and additional experimental tests are required in establishing when folding rate constants deviate from two-state behavior and a folding intermediate truly exists. In some systems, for example, the roll-over behavior can be attributed to protein aggregation (Silow and Oliveberg, 1999), movement of the transition state towards the unfolded state (Otzen et. al., 1999) or inadequate pH control upon buffer dilution. Another explanation to the deviation from the linearity in a chevron plot, obtained by using the stopped flow methodology, is related to a dead-time artifact. The saturation of the apparent rate constants, when they are on the order of the instrumental dead-time, may lead to an underestimation of the folding rate and extrapolation produces a missing burst phase amplitude even when one does not actually exist. In this case, ultrafast mixing experiments performed by continuous flow, which can monitor reactions on the microsecond time scale, may overcome this kind of limitations. The continuous flow methodology has been largely used to monitor the formation of intermediate states directly (Park et. al., 1999; Shastry et. al., 1998) here, on the contrary, this technique allowed us to disprove the existence of a folding intermediate, previously detected. By integrating stopped- and continuous-flow mixing experiments, in fact, we obtained a completely linear chevron plot, typically diagnostic of a simple two-state folding model. This behavior indicates that the burst phase in the amplitude analysis observed previously was due to an underestimation of the

observed rate constants because a large fraction of the folding reaction was lost in the dead-time of the instrument and could not be observed.

5.3.2 Folding and Binding characterization of Intrinsically disordered protein c-Myb to its binding partner KIX:

The mechanism of recognition between IDPs and their physiological partners is, in theory, a complex reaction that involves (i) the productive encounter between the two partners, (ii) the folding of the IDP system and (iii) the locking of key stabilizing interactions (Dunker et. al., 2008; Dyson et. al., 2002; Tompa, 2011; Uversky et. al., 2010; Wright and Dyson, 1999). Despite the frequent occurrence of disorder in the proteome, its role is still shrouded in mystery, and even the order of these three key events in a folding-and-binding reaction is still unclear (Kiefhaber et. al., 2012). Hagen and co-workers recently provided a complete characterization of the binding kinetics of the IDP protein IA3 to the yeast aspartic proteinase A (Narayanan et. al., 2008). It was found that IA3 binds in a disordered conformation, with a recognition event that precedes folding. On the other hand, Onitsuka et. al., investigated the folding and binding reaction of disordered mutants of staphylococcal nuclease, which fold upon ligand binding (Onitsuka et. al., 2008). Interestingly, different variants seemed to follow either the folding-after-binding or the folding-before-binding mechanism, suggesting that the energetic partitioning between the different paths shown in Scheme 1 may be fine tuned via mutagenesis, as implicit in a squared model.
In a seminal work, Wright and co-workers described the structural features of the interaction between the KIX domain and the IDP system pKID (Sugase et. al, 2008). Remarkably, by equilibrium NMR, it was found that partially folded species of pKID may bind to KIX, suggesting that, also in this case, the most plausible mechanism of IDP recognition implies a binding event preceding folding. Their equilibrium methodology, however, while providing the finest structural details of the interaction, could not unequivocally establish whether the operative recognition mechanism was a folding-after-binding rather than a folding-before-binding, a task that demands kinetic experiments. Identification of an intermediate by equilibrium measurements is not diagnostic on its role in the reaction mechanism, since it may be either on- or off- the productive pathway (Fersht Book 1999).
We present hereby an extensive kinetic characterization of the recognition reaction between KIX and c-Myb. Despite the inherent complexity of binding induced folding, it appears that, under all conditions explored, folding and binding are coupled and highly co-operative. Indeed no transient

intermediates could be detected by fluorescence spectroscopy, the time course of the T-jump induced relaxation being described by a single exponential and likewise the dissociation time course obtained by displacement was, under all conditions, consistent with a single exponential process. A comparison between the data recorded in the presence and in the absence of TFE, which stabilizes the helical conformation of c-Myb, sheds light on the order of events in the ligand induced folding. In fact, because TFE affects only the dissociation rate constants, without influence on the association rate constant, it may be concluded that in the rate limiting step c-Myb is essentially unstructured. Therefore, binding precedes folding.
It is of interest to analyze the pH dependence of the observed association and dissociation rate constants reported in Figure 4.3.9. In fact, both rate constants displayed a sigmoidal dependence on pH, each one being consistent with the titration a single group with pKa of 4.2 and 7.6 respectively.
This finding would suggest the mechanism of recognition between KIX and different IDPs to show some conserved features. Future work based on Φ-value analysis will shed light on the atomistic details of the folding and binding reactions involved in the recognition and complex formation between KIX and its different IDP partners.

REFERENCES

Abkevich, V. I., A. M. Gutin, and E. I. Shakhnovich
1994 Specific nucleus as the transition state for protein folding: evidence from the lattice model. Biochemistry 33(33):10026-36.

Adhikary, S., Eilers, M.
2005 Transcriptional regulation and transformation by Myc proteins. Nat Rev Mol Cell Biol. 6(8): 635-45.

Alexander, P. A., et al.
2007 The design and characterization of two proteins with 88% sequence identity but different structure and function. Proc Natl Acad Sci U S A 104(29):11963-8.

Alexander, P. A., et al.
2009 A minimal sequence code for switching protein structure and function. Proc Natl Acad Sci U S A 106(50):21149-54.

Alexander, P., J. Orban, and P. Bryan
1992 Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. Biochemistry 31(32):7243-8.

Alonso, D. O., and V. Daggett
1998 Molecular dynamics simulations of hydrophobic collapse of ubiquitin. Protein Sci 7(4):860-74.

Anfinsen, C. B., et al.
1961 The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc Natl Acad Sci U S A 47:1309-14.

Antonini, E.; Brunori, M
1971 *North-Holland Publ Co. Amsterdam, The Netherlands*.

Bachmann, A., and T. Kiefhaber
2001 Apparent two-state tendamistat folding is a sequential process along a defined route. J Mol Biol 306(2):375-86.

Bai, Y., et al.
1995 Protein folding intermediates: native-state hydrogen exchange. Science 269(5221):192-7.

Baker, D.
2000 A surprising simplicity to protein folding. Nature 405(6782):39-42.

Baldwin, C. T., A. M. Reginato, and D. J. Prockop
1989 A new epidermal growth factor-like domain in the human core protein for the large cartilage-specific proteoglycan. Evidence for alternative splicing of the domain. J Biol Chem 264(27):15747-50.
Baldwin, R. L.
1989 How does protein folding get started? Trends Biochem Sci 14(7):291-4.
Banachewicz, W., et al.
2011 Malleability of folding intermediates in the homeodomain superfamily. Proc Natl Acad Sci U S A 108(14):5596-601.
Bond, C. J., et al.
1997 Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: description of the folding pathway. Proc Natl Acad Sci U S A 94(25):13409-13.
Bowler, B. E.
2007 Thermodynamics of protein denatured states. Mol Biosyst 3(2):88-99.
Bowler, B. E.
2011 Residual structure in unfolded proteins. Curr Opin Struct Biol 22(1):4-13.
Brockwell, D. J., and S. E. Radford
2007 Intermediates: ubiquitous species on folding energy landscapes? Curr Opin Struct Biol 17(1):30-7.
Brunori, M., et al.
2003 Cytochrome c(551) as a model system for protein folding. Biophys Chem 100(1-3):409-19.
Brunori, M., et al.
2012 Morphogenesis of a protein: folding pathways and the energy landscape. Biochemical Society Transactions 40(2):429-32.
Bryngelson, J. D., et al.
1995 Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins 21(3):167-95.
Buchner J. and Kiefhaber J.
2005 Protein Folding Handbook, Part I. Wiley Weinheim.
Calosci, N., et al.
2008 Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. Proc Natl Acad Sci U S A 105(49):19241-6.
Capaldi, A. P., C. Kleanthous, and S. E. Radford
2002 Im7 folding mechanism: misfolding on a path to the native state. Nat Struct Biol 9(3):209-16.

Capaldi, A. P., et al.
2001 Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. Nat Struct Biol 8(1):68-72.
Cellmer, T., et al.
2007 Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration. J Am Chem Soc 129(47):14564-5.
Chattopadhyay, K., E. L. Elson, and C. Frieden
2005 The kinetics of conformational fluctuations in an unfolded protein measured by fluorescence methods. Proc Natl Acad Sci U S A 102(7):2385-9.
Chi, C.N., et al.
2007 A conserved folding mechanism for PDZ domains. FEBS Lett. 581:1109-1113.
Chiti, F., et al.
1999a Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat Struct Biol 6(11):1005-9.
Chiti, F., et al.
1999b Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. Nat. Struct. Biol. 6:1005-1009.
Cho, J. H., and D. P. Raleigh
2009 Experimental characterization of the denatured state ensemble of proteins. Methods Mol Biol 490:339-51.
Chung, H. S., J. M. Louis, and W. A. Eaton
2009 Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. Proc Natl Acad Sci U S A 106(29):11837-44.
Chung, H. S., J. M. Louis, and W. A. Eaton
2010 Distinguishing between protein dynamics and dye photophysics in single-molecule FRET experiments. Biophys J 98(4):696-706.
Clarke, J., et al.
1999a Folding studies of Ig-like beta-sandwich proteins suggest they share a common folding pathway. Structure 7:1145-1153.
Clarke, J., et al.
1999b Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. Structure 7(9):1145-53.
Colon, W., et al.

1996 Side chain packing of the N- and C-terminal helices plays a critical role in the kinetics of cytochrome c folding. Biochemistry 35(17):5538-49.

Daggett, V., and A. R. Fersht
2003 Is there a unifying mechanism for protein folding? Trends Biochem Sci 28(1):18-25.

Dalal, S., S. Balasubramanian, and L. Regan
1997 Protein alchemy: changing beta-sheet into alpha-helix. Nat. Struct. Biol. 4:548-552.

Dang, C. V.
1999 c-Myc target genes involved in cell growth, apoptosis, and metabolism. Mol Cell Biol. 19 (1): 1-11.

Davidson, A.R.
2008 A folding space odyssey. Proc. Natl. Acad. Sci. U S A 105:2759-2760.

Day, R., and V. Daggett
2003 All-atom simulations of protein folding and unfolding. Adv Protein Chem 66:373-403.

Dill, K. A.
1985 Theory for the folding and stability of globular proteins. Biochemistry 24(6):1501-9.

Dill, K. A., et al.
1995 Principles of protein folding--a perspective from simple exact models. Protein Sci 4(4):561-602.

Dill, K. A., and D. Shortle
1991 Denatured states of proteins. Annu Rev Biochem 60:795-825.

Ding, K., J. M. Louis, and A. M. Gronenborn
2004 Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR. J Mol Biol 335(5):1299-307.

Dunker, A. K., et al.
2000 Intrinsic protein disorder in complete genomes. Genome Inform Ser Workshop Genome Inform 11, 161-71.

Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L
2008 Function and structure of inherently disordered proteins. *Curr Opin Struct Biol 18*, 756-764.

Dyson, H. J., and P. E. Wright
2002 Insights into the structure and dynamics of unfolded proteins from nuclear magnetic resonance. Adv Protein Chem 62:311-40.

Dyson, H. J., and P. E. Wright
2004 Unfolded proteins and protein folding studied by NMR. Chem Rev 104(8):3607-22.

Dyson, H. J., and P. E Wright.
2005 Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6 : 197-208.
Elove, G. A., A. K. Bhuyan, and H. Roder
1994 Kinetic mechanism of cytochrome c folding: involvement of the heme and its ligands. Biochemistry 33(22):6925-35.
Eyring, H.
1935 J. Chem. Phys. 3:107.
Fanning, A. S., and J. M. Anderson
1999 PDZ domains: fundamental building blocks in the organization of protein complexes at the plasma membrane. J Clin Invest 103(6):767-72.
Ferguson, N., et al.
1999 Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. J Mol Biol 286(5):1597-608.
Fersht, A. R., A. Matouschek, and L. Serrano
1992 The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224(3):771-82.
Fersht, A. R.
1995 Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. Proc Natl Acad Sci U S A 92(24):10869-73.
Fersht, A. R.
1997 Nucleation mechanisms in protein folding. Curr Opin Struct Biol 7(1):3-9.
Fersht, A.
1999 Structure and Mechanism in Protein Science. Freeman W.H. ti
Friel, C.T., A.P. Capaldi, and S.E. Radford
2003 Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. J. Mol. Biol. (326):293-305.
Frauenfelder H., Sligar S.G., Wolynes P.G.
1991 The energy landscapes and motions of proteins. Science. Dec 13;254(5038):1598-603.
Geierhaas, C.D., et al.
2007 BPPred: a Web-based computational tool for predicting biophysical parameters of proteins. Protein Sci. 16:125-134.
Gianni , Mayor , and Fersht

2003 Structural insights in the folding of small single-domain proteins. The Italian Journal of Biochemistry 52(4).

Gianni, S

2012 THE Φ VALUE ANALYSIS. Encyclopedia of Biophysics.

Gianni, S., et al.

2009 Distinguishing between smooth and rough free energy barriers in protein folding. Biochemistry 48(49):11825-30.

Gianni, S., M. Brunori, and C. Travaglini-Allocatelli

2001a Refolding kinetics of cytochrome c(551) reveals a mechanistic difference between urea and guanidine. Protein Sci 10(8):1685-8.

Gianni, S., et al.

2005 Kinetic folding mechanism of PDZ2 from PTP-BL. Protein Eng Des Sel 18(8):389-95.

Gianni, S., et al.

2007a A PDZ domain recapitulates a unifying mechanism for protein folding. Proc Natl Acad Sci U S A 104(1):128-33.

Gianni, S., et al.

2003 Unifying features in protein-folding mechanisms. Proc Natl Acad Sci U S A 100(23):13286-91.

Gianni, S., et al.

2010 Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. Nat Struct Mol Biol 17(12):1431-7.

Gianni, S., et al.

2007b Identification and characterization of protein folding intermediates. Biophys Chem 128(2-3):105-13.

Gianni, S., et al.

2001b Snapshots of protein folding. A study on the multiple transition state pathway of cytochrome c(551) from Pseudomonas aeruginosa. J Mol Biol 309(5):1177-87.

Giri, R.,Morrone, A. Travaglini-Allocatelli, C. ,Jemth, P. ,Brunori, M.,
Gianni, S.

2012 Folding pathways of proteins with increasing degree of sequence identities but different structure and function. Proc. Natl. Acad. Sci. U S A 44, 17772-6.

Haglund E et al.

2008 Changes of protein-folding pathways by circular permutation: overlapping nuclei promote global cooperativity. J. Biol. Chem. 283:27904-27915.

Harrison, S. C., and R. Durbin

1985 Is there a single pathway for the folding of a polypeptide chain? Proc Natl Acad Sci U S A 82(12):4028-30.

He, Y., et al.
2008 NMR structures of two designed proteins with high sequence identity but different fold and function. Proc. Natl. Acad. Sci. U S A 105:14412-14417.

Horng, J.C., Tracz, S.M., Lumb, K.J., and Raleigh, D.P
(2005) Slow folding of a three-
helix protein via a compact intermediate. Biochemistry 44, 627.

Hyeon, C., and D. Thirumalai
2003 Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? Proc Natl Acad Sci U S A 100(18):10249-53.

I.E. Sanchez, T. Kiefhaber
2003 J. Mol. Biol. 327:867–884.

Itzhaki, L. S., D. E. Otzen, and A. R. Fersht
1995 The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol 254(2):260-88.

Ivarsson, Y., et al.
2008 Folding and misfolding in a naturally occurring circularly permuted PDZ domain. J Biol Chem 283(14):8954-60.

J. N. Onuchic, N. D. Socci, Z. Luthey-Schulten and P. G. Wolynes,
1996 Protein Folding Funnels: The Nature of the Transition State Ensemble. Folding & Design 1(6):441-450.

J.N. Onuchic, P.G. Wolynes
2004 Theory of Protein Folding. Curr. Opinion Structural Biology 14:70-75.

Jackson, S. E., and A. R. Fersht
1991 Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. Biochemistry 30(43):10436-43.

Jackson, S.E, Fersht, A.R.
1991 Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30:10428-10435.

Jackson, S.E.
1998 How do small single-domain proteins fold? Fold. Des. 3:R81-91.

Jasanoff, A.; Fersht, A. R. *Biochemistry*

1994 Quantitative determination of helical propensities from trifluoroethanol titration curves. *33*, 2126-2135.

Jaya, N., Garcia, V.,Vierling, E.

2009 Substrate binding site flexibility of the small heat shock protein molecular chaperones. Proc Natl Acad Sci U S A. 106 (37):15604-9.

Jemth, P., et al.

2004 Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. Proc Natl Acad Sci U S A 101(17):6450-5.

Karplus, M., and D. L. Weaver

1976 Protein-folding dynamics. Nature 260(5550):404-6.

Karplus, M. and McCammon J.A.

2002 Molecular dynamics simulations of biomolecules. Nat Struct Biol. 9(9):646-52.

Kazmirski, S. L., and V. Daggett

1998a Non-native interactions in protein folding intermediates: molecular dynamics simulations of hen lysozyme. J Mol Biol 284(3):793-806.

Kazmirski, S. L., and V. Daggett

1998b Simulations of the structural and dynamical properties of denatured proteins: the "molten coil" state of bovine pancreatic trypsin inhibitor. J Mol Biol 277(2):487-506.

Kennedy, D., and C. Norman

2005 What don't we know? Science 309(5731):75.

Kiefhaber, T.,Bachmann, A.,Jensen, K. S.

2012 Dynamics and mechanisms of coupled protein folding and binding reactions. Curr Opin Struct Biol 22 (1): 21-9.

Khorasanizadeh, S., I. D. Peters, and H. Roder

1996 Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nat Struct Biol 3(2):193-205.

Kmiecik, S., and A. Kolinski

2008 Folding pathway of the b1 domain of protein G explored by multiscale modeling. Biophys J 94(3):726-36.

Krantz, B. A., et al.

2002 Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. J Mol Biol 324(2):359-71.

Kriwacki, R. et. al.

1996 Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. Proc Natl Acad Sci U S A 93, 11504-9.

Leffler, J. E.
1953 Parameters for the Description of Transition States. *Science 117*, 340-341.
Levinthal, C.
1968 Are there pathways for protein folding? Journal de Chimie Physique et de Physico-Chimie Biologique 65:44-45.
Lindberg, M. O., and M. Oliveberg
2007 Malleability of protein folding pathways: a simple reason for complex behaviour. Curr Opin Struct Biol 17(1):21-9.
Martinez, J. C., M. T. Pisabarro, and L. Serrano
1998 Obligatory steps in protein folding and the conformational diversity of the transition state. Nat Struct Biol 5(8):721-9.
Martínez, J.C., and L. Serrano
1999 The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. Nat. Struct. Biol 6:1010-1016.
Matouschek, A., et al.
1990 Transient folding intermediates characterized by protein engineering. Nature 346(6283):440-5.
Matouschek, A., et al.
1989 Mapping the transition state and pathway of protein folding by protein engineering. Nature 340(6229):122-6.
Matthews , CR
1993 Pathways of protein folding. Annu Rev Biochem 62:653-683.
Mayor, U., et al.
2003 The complete folding pathway of a protein from nanoseconds to microseconds. Nature 421:863-867.
McCallister, E. L., E. Alm, and D. Baker
2000 Critical role of beta-hairpin formation in protein G folding. Nat Struct Biol 7(8):669-73.
Meng, W., et al.
2009 Native like structure in the unfolded state of the villin headpiece helical subdomain, an ultrafast folding protein. Protein Sci 18(8):1692-701.
Mirny, L. A., V. I. Abkevich, and E. I. Shakhnovich
1998 How evolution makes proteins fold quickly. Proc Natl Acad Sci U S A 95(9):4976-81.
Midic, U.,et. al.
2009 Protein disorder in the human diseasome: unfoldomics of human genetic diseases. BMC Genomics. 10 Suppl 1: S12.
Monod J, Wyman J, Changeux JP.

1965 On the nature of allosteric transitions: a plausible model. J Mol Biol.12:88-118.

Morrone, A. , et al.
2011a Gb1 Is not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate. . Biophysical Journal 101:1-8.

Morrone, A., et al.
2011b The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. J Biol Chem 286(5):3863-72.

Morrone, A.,Giri, R.,Brunori, M.,Gianni, S.
2012 Reassessing the folding of the KIX domain: Evidence for a two-state mechanism. Protein Sci. 21 (21).: 1775-9.

Munoz, V., and L. Serrano
1997 Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. Biopolymers 41(5):495-509.

Myers, J. K.; Pace, C. N.; Scholtz, J. M
1998 Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1. *Protein Sci 7*, 383-388.

Myers, J. K., C. N. Pace, and J. M. Scholtz
1995 Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. Protein Sci 4(10):2138-48.

Narayanan, R.; Ganesh, O. K.; Edison, A. S.; Hagen, S. J
2008 Kinetics of folding and binding of an intrinsically disordered protein: The inhibitor of yeast aspartic proteinase YPrA. *J. Am. Chem. Soc.* 2008, *130*, 11477-11485.

Neuweiler, H., C. M. Johnson, and A. R. Fersht
2009 Direct observation of ultrafast folding and denatured state dynamics in single protein molecules. Proc Natl Acad Sci U S A 106(44):18569-74.

Oldfield, S.,et. al.
2008 C-terminal splice variants of the mu-opioid receptor: existence, distribution and functional characteristics. J Neurochem. 104 (4), 937-45.

Oliveberg, M.
1998 Alternative explanations for multi-state kinetics in protein folding:

transient aggregation and changing transition-state ensambles. Acc. Chem. Res. 31:765-772.

Oliveberg, M., and A. R. Fersht
1996 Formation of electrostatic interactions on the protein-folding pathway. Biochemistry 35(8):2726-37.

Oliveberg, M., et al.
1998 The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. J Mol Biol 277(4):933-43.

Onitsuka, M., et. al.
2008 Mechanism of induced folding: both folding before binding and binding before folding can be realized in staphylococcal nuclease mutants.*Proteins*, *72*: 837-847.

Otzen, D. E., and A. R. Fersht
1998 Folding of circular and permuted chymotrypsin inhibitor 2: retention of the folding nucleus. Biochemistry 37(22):8139-46.

Otzen, D. E., et al.
1999 Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. Biochemistry 38(20):6499-511.

Ozkan, S. B., I. Bahar, and K. A. Dill
2001 Transition states and the meaning of Phi-values in protein folding kinetics. Nat Struct Biol 8(9):765-9.

Pace, C. N.
1986 Determination and analysis of urea and guanidine hydro-chloride denaturation curve. Methods Enzymol 131:266-280.

Pace, C.N.
1990 Measuring and increasing protein stability. . Trends Biotechnol. 8:93–98.

Park, S. H., K. T. O'Neil, and H. Roder
1997 An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. Biochemistry 36(47):14277-83.

Park, S. H., M. C. Shastry, and H. Roder
1999 Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. Nat Struct Biol 6(10):943-7.

Parker, M.J., J. Spencer, and A.R. Clarke
1995 An integrated kinetic analysis of intermediates and transition states in protein folding reactions. J. Mol. Biol. 253(5):771-86.

Pletneva, E. V., H. B. Gray, and J. R. Winkler

2005 Many faces of the unfolded state: conformational heterogeneity in denatured yeast cytochrome C. J Mol Biol 345(4):855-67.

Qi, P. X., Sosnick, T. R., and Englander, S. W

1998 The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. Nat Struct Biol 5: 882.

Radhakrishnan, I.et. al.

1997 Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell. 91 (6): 741-52.

Religa, T.L., et al.

2005 Solution structure of a protein denatured state and folding intermediate. Nature 437:1053-1056.

Reichmann, D.et. al.

2012 Order out of disorder: working cycle of an intrinsically unfolded chaperone. Cell, 148 (5): 947-57.

Riddle, D.S., et al.

1999 Experiment and theory highlight role of native state topology in SH3 folding. Nat. Struct. Biol. 6:1016-1024.

Romero, P et. al.

1998 Thousands of proteins likely to have long disordered regions. Pac Symp Biocomput. 437-48.

Rose, G.D., and T.P. Creamer

1994 Protein folding: predicting predicting. Proteins. Struct. Funct. Bioinf. 19:1-3.

Ruan, B., et al.

2004 Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification. Biochemistry 43:14539-14546.

Sanchez, I.E., and T. Kiefhaber

2003 Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. J. Mol. Biol. 325(2):367-376.

Santoro, M. M., and D. W. Bolen

1988 Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. Biochemistry 27(21):8063-8.

Sauder J.M., MacKenzie N.E. and Roder H.

1996 Kinetic Mechanism of Folding and Unfolding of Rhodobacter capsulatus Cytochrome c2. Biochemistry 35:16852-16862.

Scaloni, F., et al.

2010 Deciphering the folding transition state structure and denatured state properties of nucleophosmin C-terminal domain. Proc Natl Acad Sci U S A 107(12):5447-52.

Scaloni, F., et al.

2009 Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. FASEB J 23(8):2360-5.

Schellman, J. A.

1978 Solvent denaturation Biopolymers 17:1305–1322.

Scott, K. A., and J. Clarke

2005 Spectrin R16: broad energy barrier or sequential transition states? Protein Sci 14(6):1617-29.

Scott, K. A., L. G. Randles, and J. Clarke

2004 The folding of spectrin domains II: phi-value analysis of R16. J Mol Biol 344(1):207-21.

Shastry, M. C., S. D. Luck, and H. Roder

1998 A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. Biophys J 74(5):2714-21.

Shea, J. E., J. N. Onuchic, and C. L. Brooks, 3rd

2002 Probing the folding free energy landscape of the Src-SH3 protein domain. Proc Natl Acad Sci U S A 99(25):16064-8.

Shortle, D.

1995 Staphylococcal nuclease: a showcase of m-value effects. Adv Protein Chem 46:217-47.

Shortle, D., and M. S. Ackerman

2001 Persistence of native-like topology in a denatured protein in 8 M urea. Science 293(5529):487-9.

Shortle, D., and A. K. Meeker

1986 Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. Proteins 1(1):81-9.

Shoemaker, B. A., Portman, J. J., Wolynes, P. G.

2000 Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A., 97 (16): 8868-73.

Silow, M., and M. Oliveberg

1997 Transient aggregates in protein folding are easily mistaken for folding intermediates. Proc Natl Acad Sci U S A 94(12):6084-6.

Spolar, R. S.,Record, M. T., Jr.

1994 Coupling of local folding to site-specific binding of proteins to DNA. Science, 263: 777-84.

Sugase, K.; Dyson, H. J.; Wright, P. E

2007 Mechanism of coupled folding and binding of an intrinsically

disordered protein. *Nature*, *447*:1021-1025.

Syme, C. D. et. al.
2002 A Raman optical activity study of rheomorphism in caseins, synucleins and tau. New insight into the structure and behaviour of natively unfolded proteins. Eur J Biochem. 269 (1):148-56.

Szilagyi, A., et. al.
2007 Protein Folding (Handbook of neurochemistry and molecular neurobiology, Cap 10). Springer-Verlag Berlin Heidelberg.

Tanford, C.
1968 Protein denaturation. Adv Protein Chem 23:121-282.

Tapley, T. L et. al.
2009 Structural plasticity of an acid-activated chaperone allows promiscuous substrate binding. Proc Natl Acad Sci U S A. 106 (14):5557-62.

Ternstrom, T., et al.
1999 From snapshot to movie: phi analysis of protein folding transition states taken one step further. Proc Natl Acad Sci U S A 96(26):14854-9.

Timasheff, S.N.
1993 The control of protein stability and association by weak interactions with water: how do solvents affect these processes? Annu Rev Biophys Biomol Struct. 22:67-97.

Tirado-Rives, J., and W. L. Jorgensen
1993 Molecular dynamics simulations of the unfolding of apomyoglobin in water. Biochemistry 32(16):4175-84.

Tompa, P.; Kovacs, D
2010 Intrinsically disordered chaperones in plants and animals. Biochem Cell Biol 88(2):167-74.

Travaglini-Allocatelli, C., et al.
1999 Folding mechanism of Pseudomonas aeruginosa cytochrome c551: role of electrostatic interactions on the hydrophobic collapse and transition state properties. J Mol Biol 289(5):1459-67.

Travaglini-Allocatelli, C., S. Gianni, and M. Brunori
2004 A common folding mechanism in the cytochrome c family. Trends Biochem Sci 29(10):535-41.

Travaglini-Allocatelli, C., et al.
2005 An obligatory intermediate in the folding pathway of cytochrome c552 from Hydrogenobacter thermophilus. J. Biol. Chem. 280:25729-25734.

Travaglini-Allocatelli, C., et al.

2003 Exploring the cytochrome c folding mechanism: cytochrome c552 from thermus thermophilus folds through an on-pathway intermediate. J Biol Chem 278(42):41136-40.

Tsai, J., M. Levitt, and D. Baker

1999 Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. J Mol Biol 291(1):215-25.

Uversky, V. N., Gillespie, J. R.,Fink, A. L.

2000 Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41(3): 415-27.

Uversky, V. N.,Oldfield, C. J.,Dunker, A. K.

2008 Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 37: 215-46.

van Gunsteren, W. F., et al.

2001 The Key to Solving the Protein-Folding Problem Lies in an Accurate Description of the Denatured State Financial support from the Schweizerischer Nationalfonds (Project no. 21-50929.97) is gratefully acknowledged. Angew Chem Int Ed Engl 40(2):351-355.

Vendruscolo, M., et al.

2001 Three key residues form a critical contact network in a protein folding transition state. Nature 409(6820):641-5.

Wang, F.,et. al.

2012 Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. Proc Natl Acad Sci U S A. 109 (16): 6078-83.

Walkenhorst, W.F., S.M. Green, and H. Roder

1997 Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. Biochemistry 36:5795-5805.

Wetlaufer, D. B.

1973 Nucleation, rapid folding, and globular intrachain regions in proteins. Proc Natl Acad Sci U S A 70(3):697-701.

White, G. W., et al.

2005 Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. J Mol Biol 350(4):757-75.

Wildegger, G., and T. Kiefhaber

1997 Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. Journal of Molecular Biology 270(2):294-304.

Williams, R. et. al.

2001 The protein non-folding problem: amino acid determinants of intrinsic order and disorder. Pac Symp Biocomput: 89-100.

Wilson, C.A., J. Kreychman, and M. Gerstein
2000 Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. J. Mol. Biol. 297:233-249.
Wong, K. B., et al.
2000 Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. J Mol Biol 296(5):1257-82.
Wright, P. E.; Dyson, H. J
1999 Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 1999, *293*.
Wright, C. F., et al.
2003 Parallel protein-unfolding pathways revealed and mapped. Nat Struct Biol 10(8):658-62.
Zarrine-Afsar, A., S. M. Larson, and A. R. Davidson
2005 The family feud: do proteins with similar structures fold via the same pathway? Curr Opin Struct Biol 15(1):42-9.
Zarrine-Afsar, A., S. L. Lin, and P. Neudecker
2010 Mutational investigation of protein folding transition states by Phi-value analysis and beyond: lessons from SH3 domain folding. Biochem Cell Biol 88(2):231-8.
Zor, T.,De Guzman, R. N.,Dyson, H. J.,Wright, P. E.
2004 Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. J Mol Biol 337 (3): 521-34.

Biophysical Journal Volume 101 October 2011 2053–2060

GB1 Is Not a Two-State Folder: Identification and Characterization of an On-Pathway Intermediate

Angela Morrone,[†△] Rajanish Giri,[†△] Rudesh D. Toofanny,[‡] Carlo Travaglini-Allocatelli,[†] Maurizio Brunori,[†] Valerie Daggett,[‡*] and Stefano Gianni[†*]

[†]Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Università di Roma "La Sapienza", Rome, Italy; and [‡]Department of Bioengineering, University of Washington, Seattle, Washington

ABSTRACT The folding pathway of the small α/β protein GB1 has been extensively studied during the past two decades using both theoretical and experimental approaches. These studies provided a consensus view that the protein folds in a two-state manner. Here, we reassessed the folding of GB1, both by experiments and simulations, and detected the presence of an on-pathway intermediate. This intermediate has eluded earlier experimental characterization and is distinct from the collapsed state previously identified using ultrarapid mixing. Failure to identify the presence of an intermediate affects some of the conclusions that have been drawn for GB1, a popular model for protein folding studies.

INTRODUCTION

Understanding the role and structure of partially folded intermediates is of fundamental mechanistic importance for protein folding studies. Earlier work suggested that the folding of small single domain proteins generally conforms to an all-or-none behavior (1), involving simultaneous formation of secondary and tertiary structure (2). Following this view, folding occurs in a two-state fashion, via condensation around a marginally stable nucleus, and intermediates tend to be avoided (3). When the inherent stability of folding nuclei is increased, however, even very simple protein systems appear to fold in a more complex fashion (4), with population of partially folded intermediates, which may either transiently accumulate leading to multiphasic kinetics, or be a high energy species en route to the native state (5). The presence of such local minima in the landscape is very difficult to address experimentally (6) and intermediates may sometimes escape detection.

The B1 IgG-binding domain of streptococcal protein G, generally called GB1, has played a central role in protein folding studies being the system of choice in more than 200 publications carried out using a wide variety of experimental and theoretical approaches; see for example (7–19). Because of its small size and its simple and highly symmetrical topology, this small protein domain has represented an ideal candidate for a vast number of different studies. Over-and-above contrasting views on the presence of a low energy early collapsed state as detected by ultrarapid mixing (13,15,16), experimental work has been taken to indicate that GB1 folds in a canonical two-state process (20), via a polarized folding transition state with native-like structure localized in the C-terminal β-hairpin (14). On the other hand, despite the evidence arising from experiments, some theoreticians have predicted the presence of intermediates and heterogeneous pathways for the folding of GB1 (19), raising some doubts about the applicability of a bona fide two-state mechanism. In this study, we have undertaken an extensive characterization of the folding of GB1 by experiments and simulations that provide evidence for a folding intermediate. This partially structured state is an on-pathway species that, despite two decades of studies, escaped detection and characterization.

MATERIALS AND METHODS

The buffers used were 50 mM Glycine/NaOH from pH 9.6 to 9.0, 50 mM Tris/HCl from pH 9.0 to 7.2, 50 mM Bis-Tris/HCl from pH 7.0 to 6.0, 50 mM sodium acetate from pH 5.5 to 3.8, 50 mM sodium formate from pH 3.4 to 3.0, and 50 mM sodium phosphate/phosphoric acid from pH 2.8 to 2.0. All reagents were of analytical grade.

Protein expression and purification

GB1 gene was cloned into the vector pG58, which encodes an engineered subtilisin prosequence as the N-terminus of the fusion protein. Proteins were purified employing an affinity chromatography previously developed (21). Soluble cell extract of prodomain fusion proteins was injected onto a 5-ml Bio-Scale Mini Profinity eXact cartridge (Bio-Rad, Hercules, CA) at 5 ml/min to allow binding and then washed with 10-column volumes of 100 mM $NaPO_4$ (pH 7.2) to remove impurities. To cleave and elute the purified target proteins, 15 ml of 100 mM NaF in the presence of 100 mM $NaPO_4$ (pH 7.2) were injected at 5 ml/min. After the first 10 ml, the flow was stopped and the column incubated for 30 min to allow complete cleavage. After elution, the purified proteins were then dialyzed to remove sodium fluoride.

Equilibrium unfolding

Equilibrium denaturations were carried out on a Fluoromax single photon counting spectrofluorometer (Jobin-Yvon, Edison, NJ). Tryptophan

Submitted July 22, 2011, and accepted for publication August 19, 2011.

△Angela Morrone and Rajanish Giri contributed equally to this work.

*Correspondence: daggett@uw.edu or stefano.gianni@uniroma1.it

This article is dedicated to the late Professor Quentin H. Gibson, the inventor of the stopped flow spectrophotometer in 1954, who passed away on March 16, 2011.

Editor: Doug Barrick.

0006-3495/11/10/2053/8 $2.00

doi: 10.1016/j.bpj.2011.09.013

fluorescence emission spectra were recorded in a cuvette (1-cm light path) between 300 and 400 nm. The excitation wavelength was 280 nm. Protein concentrations were typically 3 μM.

Stopped-flow measurements

Single mixing kinetic folding experiments were carried out on a Pi-star or on an SX-18 stopped-flow instrument (Applied Photophysics, Leatherhead, UK); the excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. In all experiments, performed at 25°C, refolding and unfolding were initiated by an 11-fold dilution of the denatured or the native protein with the appropriate buffer. Final protein concentrations were typically 1 μM. The observed kinetics were always independent of protein concentration (from 0.5 to 5 μM after mixing protein concentration), as expected from monomolecular reactions without effects due to transient aggregation (22).

Data analysis

Equilibrium experiments

Assuming a standard two-state model, the guanidine-induced denaturation transitions were fitted to the following equation:

$$\Delta G_d = m_{D-N}(D - D_{50}),$$

where m_{D-N} is the slope of the transition (proportional to the increase in solvent-accessible surface area on going from the native to the denatured state) and D_{50} is the midpoint of the denaturation transition. An equation, which takes into account the pre- and post-transition baselines, was used to fit the observed unfolding transition (23). Equilibrium folding stabilities calculated at different pH conditions are listed in Table 1.

Kinetic experiments

Analysis was performed by nonlinear least-squares fitting of single exponential phases using the fitting procedures provided in the Applied Photophysics software. The chevron plots were fitted globally by numerical analysis based on a three-state model as discussed in the Results section. The global fit of chevron plots at different pH was obtained with Prism software (Graphpad).

Molecular dynamics simulation methods

All-atom explicit solvent molecular dynamics (MD) simulations were performed for protein GB1 by starting with the NMR structure (PDB 3GB1 (24)). Simulations included one native state simulation at 25°C (21 ns) and five unfolding simulations at 225°C (1×41 ns, 1×38, 3×2 ns) resulting in a total of 106 ns of simulation time. All simulations were performed using our in-house MD software, *in lucem* molecular mechanics (25) with the Levitt et al. (26) all-atom force field. The microcanonical ensemble was used, where the number of atoms, unit cell volume, total energy, and linear momentum were conserved. The protein was solvated in a periodic box of flexible F3C water molecules (27). For a more in-depth discussion of the simulation protocols please see (28).

One-dimensional reaction coordinate

To identify the transition and intermediate states, we calculated a multidimensional property space derived from 15 physical properties of the protein and then calculated a one-dimensional (1D) reaction coordinate based on these 15 properties (29). The 15 properties were native contacts, nonnative contacts, radius of gyration, end-to-end distance, main-chain solvent accessible surface area (SASA) (30) side-chain SASA, polar SASA, nonpolar SASA, main-chain polar SASA, main-chain nonpolar SASA, side-chain polar, side-chain nonpolar SASA, total SASA, fraction of helix, and fraction of β-structure. The distance between two structures in property space is calculated as the average Euclidean distance between the 15-dimensional points. The mean distance in property space was calculated for each time point in a simulation of interest to the native state reference, which contained all the structures of the native state simulation excluding the first nanosecond. The 1D reaction coordinate was created from a histogram of the mean distance to reference for all structures. To compare with

TABLE 1 Folding parameters of GB1 as a function of pH

pH	k_{DI} (s^{-1})	$k_{NI}/(1 + K_{part})$ (s^{-1})	K_{part}	$\Delta G_{D\text{-}N}$* (kcal mol^{-1})	$\Delta G_{D\text{-}N}^{\dagger}$ (kcal mol^{-1})
2.0	1000 ± 90	11.0 ± 3.0	1.7 ± 0.2	2.7 ± 0.3	3.9 ± 0.8
2.5	1000 ± 90	19.2 ± 4.8	3.0 ± 0.4	2.3 ± 0.2	3.0 ± 0.7
3.0	1050 ± 95	5.2 ± 1.3	0.9 ± 0.1	3.1 ± 0.4	3.8 ± 0.4
3.5	1200 ± 120	5.1 ± 1.3	1.2 ± 0.1	3.2 ± 0.2	3.9 ± 0.4
4.0	1400 ± 120	1.3 ± 0.3	0.42 ± 0.09	4.1 ± 0.2	4.5 ± 0.4
4.5	1800 ± 170	0.4 ± 0.1	0.20 ± 0.04	5.0 ± 0.2	4.2 ± 0.4
5.0	1300 ± 120	0.15 ± 0.04	0.09 ± 0.02	5.3 ± 0.2	4.9 ± 0.5
5.5	1370 ± 100	0.15 ± 0.04	0.13 ± 0.03	5.4 ± 0.1	4.6 ± 0.5
6.0	720 ± 70	0.12 ± 0.03	0.09 ± 0.02	5.1 ± 0.1	5.2 ± 0.5
6.5	830 ± 70	0.14 ± 0.04	0.10 ± 0.02	5.1 ± 0.2	4.7 ± 0.5
7.0	670 ± 70	0.19 ± 0.05	0.12 ± 0.02	4.8 ± 0.2	5.3 ± 0.5
7.5	630 ± 70	0.27 ± 0.07	0.11 ± 0.02	4.6 ± 0.2	4.3 ± 0.4
8.0	600 ± 60	0.4 ± 0.1	0.12 ± 0.03	4.3 ± 0.3	4.3 ± 0.4
8.5	500 ± 40	0.6 ± 0.2	0.11 ± 0.02	3.9 ± 0.2	3.8 ± 0.4
9.0	390 ± 40	0.9 ± 0.2	0.13 ± 0.03	3.6 ± 0.2	3.3 ± 0.3
9.6	220 ± 25	1.3 ± 0.3	0.11 ± 0.02	3.0 ± 0.2	3.0 ± 0.3

*Calculated from chevron plot analysis. The Chevron plots were fitted globally to a three-state model with shared *m*-values. k_{DI} is the microscopic rate constant for the formation of the intermediate from the denatured state; k_{NI} is the microscopic rate constant for the unfolding of the native state to the intermediate state; K_{part} is the partitioning factor k_{ID}/k_{IN} reflecting the difference between the activation barriers for the intermediate to revert to the reagents rather than proceeding to the products. The analysis returned a total $m_{D\text{-}N} = 1.95 \pm 0.2$ kcal mol^{-1} M^{-1}. The Tanford β-values for the two transition states were $\beta_{TS1} = 0.76 \pm 0.04$ and $\beta_{TS2} = 0.93 \pm 0.04$.

†Calculated from equilibrium denaturation. Equilibrium denaturations were fitted both individually and globally with shared $m_{D\text{-}N}$ value. The global analysis returned $m_{D\text{-}N} = 1.75 \pm 0.2$ kcal mol^{-1} M^{-1}. This value was consistent within error with the values obtained by fitting individually each independent equilibrium experiments, as well as with the value calculated from kinetic experiments.

experimental Tanford β-values, we calculated the ratio of the average total (SASA) for the transition state ensemble and the average total SASA for the native state simulation.

To investigate the unfolding pathway of GB1 we calculated contact matrices, for each state identified, based on the fraction of time that the residues were in contact. A pair of residues was considered in contact if it contained carbon atoms that were <5.4 Å apart or one carbon atom or any other atoms <4.6 Å.

RESULTS

Equilibrium denaturation of GB1

To study the folding mechanism of GB1, we carried out both equilibrium and kinetic experiments. Guanidine-induced equilibrium denaturation of GB1, monitored by fluorescence spectroscopy, was investigated at 25°C exploring a wide range of pH values, from 2.0 to 9.6 (data not shown). Equilibrium denaturation curves were fitted both individually and globally with a shared $m_{\text{D-N}}$ value. Parameters obtained from global analysis, listed in Table 1, were consistent within error with the values obtained by fitting each independent equilibrium experiment individually, as well as with the values calculated from kinetic experiments, confirming the two-state nature of the equilibrium unfolding transition of GB1. The denaturation profiles were all consistent with two-state unfolding and yielded an m-value of 1.75 ± 0.2 kcal mol^{-1} M^{-1}.

Kinetic experiments

The folding and unfolding kinetics of GB1 were investigated at several pH values, ranging from 2.0 to 9.6. In all cases, folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final guanidine concentration. Each rate constant was obtained from the average of at least five independent shots in stopped-flow experiments. Semilogarithmic plots of the observed folding/unfolding rate constant of GB1 versus denaturant concentration (chevron plot) at the different pH values are presented in Fig. 1. Surprisingly, the unfolding arm of the chevron plots at pH values higher than 6.0 shows a deviation from linearity that becomes evident at high guanidine concentrations (rollover effect). This deviation is highlighted in Fig. 2 where the chevron plot of GB1 measured at pH 9.0 is reported together with the residuals of the fit, showing a clear systematic deviation from the expected values for a two-state behavior. This effect escaped previous studies probably because of the restricted range of experimental conditions, limited to [GdnHCl] < 5.5 M (14,15). Indeed, if we were to ignore the data we recorded for [GdnHCl] > 5.5 M, the unfolding arm of the chevron plots would appear essentially linear but would display a puzzling change in slope with the different experimental conditions.

Analysis of chevron plots is a common and powerful tool for detecting protein folding intermediates (31). In fact, if there is only one rate-limiting energy barrier, the logarithm

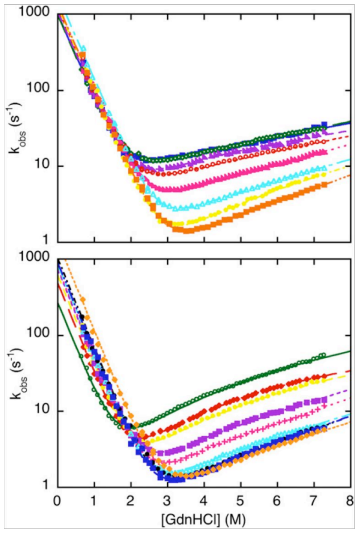


FIGURE 1 Folding kinetics of GB1. Top panel: Chevron plots measured from pH 2.0 to pH 5.5 at 25°C (*blue*, pH 2.0; *green*, pH 2.5; *violet*, pH 3.0; *red*, pH 3.5; *magenta*, pH 4.0; *cyan*, pH 4.5; *yellow*, pH 5.0; *orange*, pH 5.5). Bottom panel: Chevron plots measured from pH 5.5 to pH 9.6 at 25°C (*orange*, pH 5.5; *blue*, pH 6.0; *black*, pH 6.5; *cyan*, pH 7.0; *magenta*, pH 7.5; *violet*, pH 8.0; *yellow*, pH 8.5; *red*, pH 9.0; *green*, pH 9.6). Lines are the best global fit to a three-state equation with shared m-values. If data at [GdnHCl] > 5.5 M were to be ignored, a quasilinear unfolding arm with an apparent change in unfolding m-values would be seen.

of the observed folding and unfolding kinetics is expected to comply with a V-shaped dependence (32). Therefore, a deviation from linearity in either the folding or the unfolding branches of the curve may be considered of diagnostic value for the identification of intermediates (33,34). If a partially folded intermediate is present, the folding kinetics can be described by a three-state mechanism:

$$\text{D} \underset{k_{\text{ID}}}{\overset{k_{\text{DI}}}{\rightleftharpoons}} \text{I} \underset{k_{\text{NI}}}{\overset{k_{\text{IN}}}{\rightleftharpoons}} \text{N}, \qquad \text{(Scheme 1)}$$

where k_{DI} is the microscopic rate constant for the formation of the intermediate from the denatured state, k_{ID} is the microscopic rate constant for the unfolding of the intermediate to the denatured state, k_{IN} is the microscopic rate constant for the formation of the native state from the intermediate, and k_{NI} is the microscopic rate constant for the unfolding of the native to the intermediate state. Two approximations have been introduced to describe the folding pathway of three-state systems. The intermediate may be assumed to be in a fast preequilibrium with one of the

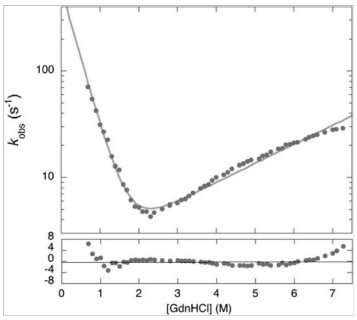


FIGURE 2 Chevron plot of wild-type GB1 measured at pH 9.0. The gray line is the best fit to a two-state equation; the residuals of the fit below show a clear systematic deviation from the expected values for a two-state behavior. Therefore, data at [GdnHCl] > 5.5 M are critical to detect the roll-over effect.

ground states, the curvature in the chevron plot being caused by accumulation of the intermediate (33). Alternatively, if the intermediate is assumed to be at steady state and at negligible concentration, the curvature in the observed (un) folding kinetics arises from a change in the rate-limiting step with changing denaturant concentration (34–38). Both approximations lead to nearly identical numerical solutions and are thus indistinguishable based only on analysis of the rate constants. In theory, however, accumulation of the intermediate should lead to multiphasic observed kinetics and more than one relaxation rate constant should be observed. In the case of GB1, because we observed single exponential folding kinetics under all conditions explored, we favored a model involving a high-energy on-pathway intermediate and the observed chevron plots were fitted to the following equation:

$$k_{\text{obs}} = k_{\text{DI}} + \frac{k_{\text{NI}}}{1 + K_{\text{part}}},$$

where K_{part} is the partition factor $k_{\text{ID}}/k_{\text{IN}}$ proportional to the difference between the activation barriers for the intermediate state to revert to the reagents rather than proceeding to the products. The logarithm of each microscopic rate constant was assumed to vary linearly with denaturant concentration (the slope of each dependence yielding the corresponding m-value). The observed chevron plots were fitted globally with shared m-values. Parameters calculated from global analysis (listed in Table 1) allow the identification of the relative positions of the two activation barriers along the reaction coordinate in terms of their relative accessible surface area (Tanford β-value), resulting in β_{T}-values of 0.76 ± 0.04 for the transition state TS1 and of 0.93 ± 0.04 for the native-like activation barrier TS2. The excellent statistical parameters of the global analysis indicate that the two activation barriers are robust to changes in pH conditions and display a conserved solvent accessible surface area when pH and protein stability are altered.

The effect of pH on the folding kinetics of GB1

The analysis of the chevron plots reported in Fig. 1 allowed determination of the folding and unfolding rate constants of GB1 over a very wide range of pH. The folding rate constants display a negligible dependence on pH (data not shown). A plot of the logarithm of the apparent unfolding rate constants from the native state to TS2 (k_{NI}) and TS1 (formally equivalent to $k_{\text{NI}}/(1 + K_{\text{part}})$) as a function of pH are reported in Fig. 3. Interestingly, both TS1 and TS2 display sigmoidal transitions at acidic and alkaline pH consistent with protonation of at least two groups in the native state with apparent pK_a values of ~4 and ~8. Importantly, the acid transition for the unfolding rate constant of TS1 (changing by almost two orders of magnitude) is more pronounced than that for the unfolding rate constant of TS2 (changing by less than an order of magnitude), suggesting the contribution of a salt bridge that is weak or not formed in TS1, but is consolidated in TS2. As detailed in the Discussion section, inspection of the three-dimensional structure of GB1 and the unfolding profiles obtained by MD simulations suggest the acidic group to be either Glu-15 or Glu-56.

MD simulations

To further explore the putative intermediate state, all-atom explicit solvent MD simulations of GB1 were carried out. Multiple simulations were performed to model thermal unfolding (at 225°C) as well as native-state behavior at 25°C (control). The transition and intermediate states were

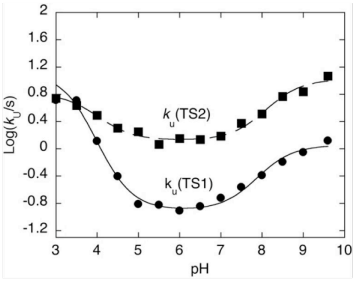


FIGURE 3 Unfolding rate constants versus pH. Logarithm of the calculated unfolding rate constants from the native state N to the first transition state k_u(TS1), and from the native state to the second transition state k_u(TS2) as a function of pH. The lines are the best fit to a model involving the protonation of two groups. In both cases, we obtained approximately the same pK_a of ~4 and ~8.

identified by calculating a multidimensional property space derived from 15 physical properties of the protein and then embedding this within a 1D reaction coordinate based on the 15 properties (as detailed in Materials and Methods) (Fig. 4). Five independent thermal unfolding simulations were performed; all simulations passed through the intermediate state (Fig. 5) however only two of them (runs 2 and 4) showed the buildup of an intermediate as defined by the 1D reaction coordinate in Fig. 4 and thus warranted further analysis. The native, transition, intermediate and denatured states were identified for the two simulations displaying buildup of the intermediate. The calculated Tanford β_T-values based on simulations, in particular on total SASA, were 0.94 and 0.92 for TS2 in simulations 2 and 4, respectively, in agreement with the experimentally derived value of 0.93 $\pm$ 0.04. For TS1 the calculated β_T-values were 0.73 and 0.75 for runs 2 and 4, respectively, in agreement with the experimental value of 0.76 $\pm$ 0.04.

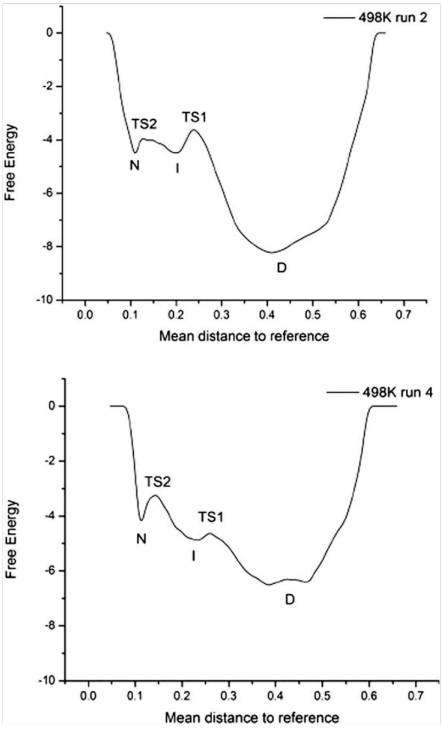


FIGURE 4 Free energy reaction coordinate for two unfolding simulations of GB1. A 1D reaction coordinate is created by calculating the mean distance to reference in a 15-dimensional property space for all structures in an unfolding simulation. The free energy reaction coordinates are calculated by taking the negative log of the counts in a mean distance to the reference histogram. Each 498 K simulation is shown with a black line. The location of unfolding state D is indicated on each free energy reaction coordinate. Top: Run 2 shows three-state unfolding with TS2 located between 0.12 and 0.13 mean distance to reference and TS1 located between 0.23 and 0.25 mean distance to reference. Bottom: Run 4 also shows three-state unfolding with TS2 located between 0.135 and 0.15 mean distance to reference and TS1 located between 0.25 and 0.265 mean distance to reference.

Considering the unfolding runs in reverse, the simulated folding pathways were largely conserved in the different runs (Fig. 6), the main difference originating from formation of nonnative contacts in the loop between strands A and B. Starting in the denatured state, nonnative contacts dominated the contact matrix and there was residual helix, which gave rise to native contacts. Considering TS1, contacts began to form between strands A and B, strands A and D, and strands C and D; the helix was nearly fully formed. The intermediate contained also contacts between strand B and the helix. There were still native contacts between strands A and B. The N-terminal ends of strand A formed contacts with strand D, although many were short-term nonnative contacts. Contacts also formed between strand C and the helix as the latter began to move toward the protein core. Contacts between strands C and D were present and almost fully formed. From I to TS2 the protein gained native contacts between the helix and strand A that allowed the helix to dock to the core. Contacts between strands A and B, strands C and D, and strands A and D became almost fully native-like. In moving from TS2 to the native state, the helix docked to the core of the

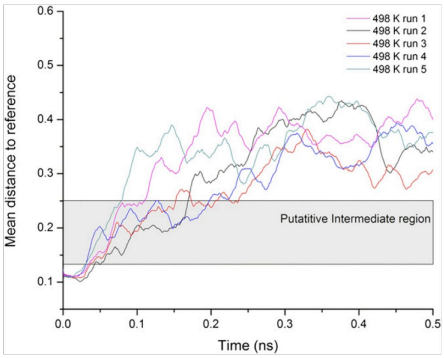


FIGURE 5 Mean distance to reference versus time for each unfolding simulation at 498 K. Lines have been smoothed using adjacent averaging to remove noise. The intermediate state region (0.135 to 0.15) in property space was determined from the positions of TS1 and TS2 in the free energy reaction coordinates from Fig. 4 for runs 2 and 4. It is clear that runs 2 and 4 populate this region significantly while the others pass through the intermediate more rapidly.

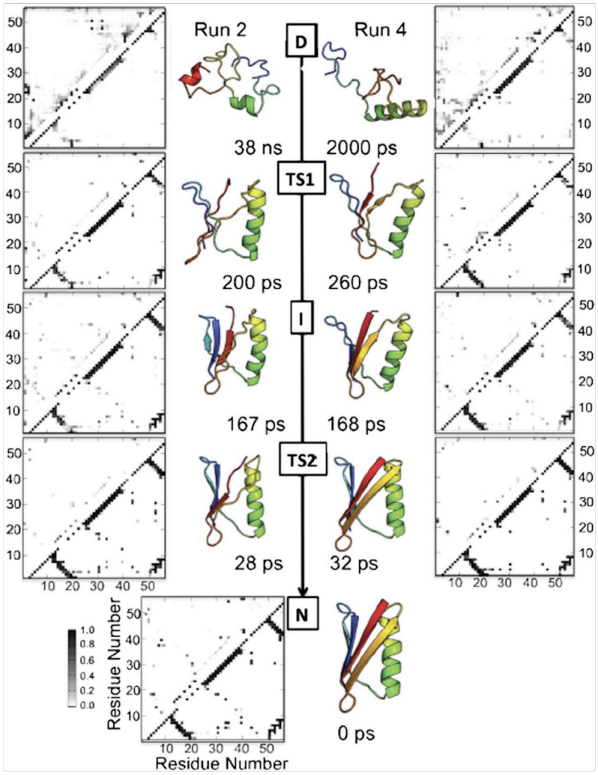


FIGURE 6 Folding pathway and contact maps for two unfolding simulations of GB1. Representative structures from each state are colored in rainbow from red to blue. The native structure of GB1 contains four β-strands (A = *red*, B = *orange*, C = *cyan*, and D = *blue*) and a helix. Contact maps show the fraction of time residues were in contact with nonnative interactions (*above the diagonal*) or with native contacts (*below*). Contacts are colored in a grayscale from white to black, where white indicates that the residues were never in contact and black those that were in contact 100% of the time. Native contacts are reported over the full 498 K simulation (n = 20,000). TS, I and D contacts are reported for the time points constituting the ensemble of structures as determined by the free energy reaction coordinate (Fig. 5).

protein making contacts with strands A, B, and C. The contacts between strands A and B, strands C and D, and strands A and D became fully formed.

DISCUSSION

We have carried out an extensive characterization of the folding pathway of GB1 by experiments and MD simulations. The kinetic experiments, carried out over a wide range of pH values (from 2.0 to 9.6), highlighted the presence of a rollover effect in the chevron plots, which is typically diagnostic of the presence of a folding intermediate. An intermediate was detected also by unfolding MD simulations, which allowed us to present a structural characterization of the (un)folding pathway.

The existence of an intermediate in the folding of GB1 has been previously proposed by Roder and co-workers (15,16), who suggested the presence of a collapsed state accumulating in the μs time range, as observed by continuous-flow ultrarapid mixing experiments. This low-energy intermediate (which later was questioned by Sosnick and co-workers (13)) was also consistent with a detailed ab initio simulation reported by Kmiecik and Kolinski (39). Importantly, however, the partially folded state identified in our work is distinct from that previously suggested, being a high energy species that never accumulates; nevertheless, its existence suggests that GB1 folds via a complex and rough energy barrier with at least two discrete major transition states. Thus, while the equilibrium unfolding is consistent with a simple two-state model (7,8,20), the folding and unfolding kinetics are more complex.

When folding is characterized by a complex chevron plot, the deviation from linearity observed may have different origins: i), the curvature may be due to a movement of the position of the transition state along a single broad barrier (40,41); or ii), it may reflect a change in the rate-limiting

step suggestive of a multistate process (34,35,37,38). In the case of GB1, the curved chevron plots are observed only above pH 6 (Fig. 1), making the broad transition state model less likely because it would imply that it is possible to drastically distort the folding free energy profile and to switch between a narrow energy maximum (linear chevron plot) to a broad energy maximum (curved chevron plot) by changing pH. Furthermore, in analogy to what has been observed previously for other proteins (35,42,43), detection of a rollover effect only under some solvent conditions seems more consistent with a multistep folding pathway. The unbiased unfolding MD simulations reported in this work, which identified a folding intermediate, also support this interpretation. The excellent statistical parameters of the global fit suggest that the two transition states surrounding the intermediate are relatively robust and maintain their overall structural features when solvent conditions are varied. This observation is in stark contrast with previous experimental work, which suggested the unfolding *m*-value to depend strongly on experimental conditions (14–16). We conclude that both TS1 and TS2 display a robust structure that is by and large maintained when solvent and/or sequence composition is altered.

It is of interest to analyze the dependence on pH of the unfolding rate constants measured for the two transition states (Fig. 3). Both energy barriers appear to display a monotonic transition at acidic and alkaline pH values. Both profiles are consistent with a model involving the protonation of two different groups with apparent pK_a values of ~4 and ~8 (Fig. 3). While in the alkaline region the total change in activation free energy is approximately the same for TS1 and TS2, the acidic transition is more pronounced in TS1, suggesting the presence of a charged interaction that is weak or not formed in TS1 but consolidated in TS2. Inspection of the three-dimensional structure of GB1 suggests that such a salt bridge may be either Lys-4-Glu-15 (located at the N-terminal β-hairpin, strands A and B) or Lys-10-Glu-56 (between the N-terminal turn and the C terminus of the protein). Both interactions would be consistent with MD simulations, suggesting that the contacts between strands A and B, strands C and D, and strands A and D are just incipient in TS1, whereas they are almost fully formed in TS2.

The thermal unfolding simulations are in good agreement with the experiments and reveal an on-pathway intermediate. The β_T-values for the transition states on either side of the intermediate ensemble are in excellent agreement with the experimentally determined ones. The structure of the intermediate is very native-like with reduced native contacts between β-strands and loss of contacts between the helix and the core of the protein, which facilitates undocking (Fig. 6). Importantly, although some studies have proposed that GB1 folds via a polarized transition state, with native-like structure located in the C-terminal hairpin (14,44), our simulations suggest folding occurs via a slightly different mechanism. In fact, structure formation appears to involve a more extended nucleus, which is stabilized by both the N- and C-terminal β-hairpins, as well as by contacts between the N- and C-terminal strands. This scenario suggests the presence of multiple overlapping folding nuclei (or a diffuse, rather than polarized, nucleus).

To conclude, we have reassessed the folding of GB1 by kinetic experiments and MD simulations and show that, while its equilibrium unfolding conforms to a two-state mechanism, folding and unfolding kinetics are more complex and involve the presence of an energy barrier with at least two discrete transition states and one on-pathway intermediate. The structures of these transition states, which we have assessed by MD simulations, appear robust to changes in pH and are characterized by an extended nucleus, which is stabilized by both the N- and C-terminal β-hairpins, as well as by contacts between the N- and C-terminal strands.

This work was supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (RBRN07BMCT_007, to M.B.) and from the National Institutes of Health (GM50789 to V.D.).

REFERENCES

1. Fersht, A. R. 2008. From the first protein structures to our current knowledge of protein folding: delights and scepticisms. *Nat. Rev. Mol. Cell Biol.* 9:650–654.
2. Itzhaki, L. S., D. E. Otzen, and A. R. Fersht. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254:260–288.
3. Fersht, A. R. 1995. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA.* 92:10869–10873.
4. White, G. W., S. Gianni, …, V. Daggett. 2005. Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J. Mol. Biol.* 350:757–775.
5. Gianni, S., N. R. Guydosh, …, A. R. Fersht. 2003. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. USA.* 100:13286–13291.
6. Bryngelson, J. D., J. N. Onuchic, …, P. G. Wolynes. 1995. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins.* 21:167–195.
7. Alexander, P., J. Orban, and P. Bryan. 1992. Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. *Biochemistry.* 31:7243–7248.
8. Chung, H. S., J. M. Louis, and W. A. Eaton. 2010. Distinguishing between protein dynamics and dye photophysics in single-molecule FRET experiments. *Biophys. J.* 98:696–706.
9. Ding, K., J. M. Louis, and A. M. Gronenborn. 2004. Insights into conformation and dynamics of protein GB1 during folding and unfolding by NMR. *J. Mol. Biol.* 335:1299–1307.
10. Frank, M. K., G. M. Clore, and A. M. Gronenborn. 1995. Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci.* 4:2605–2615.
11. Hubner, I. A., J. Shimada, and E. I. Shakhnovich. 2004. Commitment and nucleation in the protein G transition state. *J. Mol. Biol.* 336: 745–761.

12. Islam, S. A., M. Karplus, and D. L. Weaver. 2004. The role of sequence and structure in protein folding kinetics; the diffusion-collision model applied to proteins L and G. *Structure.* 12:1833–1845.
13. Krantz, B. A., L. Mayne, …, T. R. Sosnick. 2002. Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. *J. Mol. Biol.* 324:359–371.
14. McCallister, E. L., E. Alm, and D. Baker. 2000. Critical role of beta-hairpin formation in protein G folding. *Nat. Struct. Biol.* 7:669–673.
15. Park, S. H., K. T. O'Neil, and H. Roder. 1997. An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. *Biochemistry.* 36:14277–14283.
16. Park, S. H., M. C. Shastry, and H. Roder. 1999. Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. *Nat. Struct. Biol.* 6:943–947.
17. Sheinerman, F. B., and C. L. Brooks, 3rd. 1997. A molecular dynamics simulation study of segment B1 of protein G. *Proteins.* 29:193–202.
18. Shen, T., C. Zong, J. J. Portman, and P. G. Wolynes. 2008. Variationally determined free energy profiles for structural models of proteins: characteristic temperatures for folding and trapping. *J. Phys. Chem. B.* 112:6074–6082.
19. Shimada, J., and E. I. Shakhnovich. 2002. The ensemble folding kinetics of protein G from an all-atom Monte Carlo simulation. *Proc. Natl. Acad. Sci. USA.* 99:11175–11180.
20. Chung, H. S., J. M. Louis, and W. A. Eaton. 2009. Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories. *Proc. Natl. Acad. Sci. USA.* 106:11837–11844.
21. Alexander, P. A., Y. He, …, P. N. Bryan. 2007. The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc. Natl. Acad. Sci. USA.* 104:11963–11968.
22. Silow, M., and M. Oliveberg. 1997. Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl. Acad. Sci. USA.* 94:6084–6086.
23. Santoro, M. M., and D. W. Bolen. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry.* 27:8063–8068.
24. Kuszewski, J., A. M. Gronenborn, and G. M. Clore. 1999. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* 121:2337–2338.
25. Beck, D. A. C., D. O. Alonso, and V. Daggett. 2000–2011. *in lucem* Molecular Mechanics (*il*mm). University of Washington, Seattle, WA.
26. Levitt, M., M. Hirshberg, …, V. Daggett. 1995. Potential-energy function and parameters for simulations of the molecular-dynamics of proteins and nucleic-acids in solution. *Comput. Phys. Commun.* 91:215–231.
27. Levitt, M., M. Hirshberg, …, V. Daggett. 1997. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J. Phys. Chem. B.* 101:5051–5061.
28. Beck, D. A. C., and V. Daggett. 2004. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods.* 34: 112–120.
29. Toofanny, R. D., A. L. Jonsson, and V. Daggett. 2010. A comprehensive multidimensional-embedded, one-dimensional reaction coordinate for protein unfolding/folding. *Biophys. J.* 98:2671–2681.
30. Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.
31. Gianni, S., Y. Ivarsson, …, C. Travaglini-Allocatelli. 2007. Identification and characterization of protein folding intermediates. *Biophys. Chem.* 128:105–113.
32. Jackson, S. E., and A. R. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry.* 30: 10428–10435.
33. Parker, M. J., J. Spencer, and A. R. Clarke. 1995. An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J. Mol. Biol.* 253:771–786.
34. Wildegger, G., and T. Kiefhaber. 1997. Three-state model for lysozyme folding: triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.* 270:294–304.
35. Bachmann, A., and T. Kiefhaber. 2001. Apparent two-state tendamistat folding is a sequential process along a defined route. *J. Mol. Biol.* 306:375–386.
36. Sánchez, I. E., and T. Kiefhaber. 2003. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* 325:367–376.
37. Sauder, J. M., N. E. MacKenzie, and H. Roder. 1996. Kinetic mechanism of folding and unfolding of Rhodobacter capsulatus cytochrome c2. *Biochemistry.* 35:16852–16862.
38. Walkenhorst, W. F., S. M. Green, and H. Roder. 1997. Kinetic evidence for folding and unfolding intermediates in staphylococcal nuclease. *Biochemistry.* 36:5795–5805.
39. Kmiecik, S., and A. Kolinski. 2008. Folding pathway of the b1 domain of protein G explored by multiscale modeling. *Biophys. J.* 94:726–736.
40. Oliveberg, M. 1998. Alternative explanations for multi-state kinetics in protein folding: transient aggregation and changing transition-state ensambles. *Acc. Chem. Res.* 31:765–772.
41. Oliveberg, M., Y. J. Tan, …, A. R. Fersht. 1998. The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding. *J. Mol. Biol.* 277:933–943.
42. Gianni, S., M. Brunori, …, M. Zhang. 2009. Distinguishing between smooth and rough free energy barriers in protein folding. *Biochemistry.* 48:11825–11830.
43. Gianni, S., N. Calosci, …, C. Travaglini-Allocatelli. 2005. Kinetic folding mechanism of PDZ2 from PTP-BL. *Protein Eng. Des. Sel.* 18:389–395.
44. Karanicolas, J., and C. L. Brooks, 3rd. 2002. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.* 11:2351–2361.

Biochemical Society Transactions www.biochemsoctrans.org

Morphogenesis of a protein: folding pathways and the energy landscape[1]

Maurizio Brunori[2], Stefano Gianni, Rajanish Giri, Angela Morrone and Carlo Travaglini-Allocatelli
Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche 'A. Rossi Fanelli', Università di Roma 'La Sapienza', Rome, Italy

Abstract

Current knowledge on the reaction whereby a protein acquires its native three-dimensional structure was obtained by and large through characterization of the folding mechanism of simple systems. Given the multiplicity of amino acid sequences and unique folds, it is not so easy, however, to draw general rules by comparing folding pathways of different proteins. In fact, quantitative comparison may be jeopardized not only because of the vast repertoire of sequences but also in view of a multiplicity of structures of the native and denatured states. We have tackled the problem of the relationships between the sequence information and the folding pathway of a protein, using a combination of kinetics, protein engineering and computational methods, applied to relatively simple systems. Our strategy has been to investigate the folding mechanism determinants using two complementary approaches, i.e. (i) the study of members of the same family characterized by a common fold, but substantial differences in amino acid sequence, or (ii) heteromorphic pairs characterized by largely identical sequences but with different folds. We discuss some recent data on protein-folding mechanisms by presenting experiments on different members of the PDZ domain family and their circularly permuted variants. Characterization of the energetics and structures of intermediates and TSs (transition states), obtained by Φ-value analysis and restrained MD (molecular dynamics) simulations, provides a glimpse of the malleability of the dynamic states and of the role of the topology of the native states and of the denatured states in dictating folding and misfolding pathways.

Introduction

Acquisition of a well defined and functionally competent three-dimensional structure from a disordered polypeptide is a complex morphogenetic event that is basic to the life of a cell. After synthesis at the ribosome, the nascent polypeptide chain is exposed to many different stimuli and potential attacks, demanding an efficient and precise mechanism whereby the native state is quickly populated. Since for small proteins, folding occurs spontaneously, it was concluded 50 years ago that the native structure is the most stable accessible conformation, i.e. it corresponds to the lowest free-energy state [1]. It is astonishing how rapidly a polypeptide can find the correct conformation starting from a myriad of random coil structures. Given that folding cannot occur via a stochastic search among all possible conformations, the existence of a preferential pathway in the transition from the denatured to the native conformation has been classically invoked to account for the efficiency of folding [2]. Therefore the traditional approach to protein folding focused on the role of intermediates thought to be obligatory sequential states in the overall process.

Since the process whereby a protein acquires its native three-dimensional shape involves formation of many non-covalent weak bonds, a classical one-trajectory view of protein folding is likely to be an over-simplification of the underlying mechanism. An original viewpoint emerged when the concept of energy landscape was extended to folding, described as a process taking place on a rugged free-energy surface, shaped as a funnel and involving entropy–enthalpy compensation [3]. Although the presentation of the funnel model provided a novel outlook on protein folding, a detailed experimental description of such a complex scenario is still a challenge. In the present paper, we describe some recent experimental work, partly carried out in Rome, and briefly discuss its significance with reference to the energy landscape theory.

Key words: energy landscape, folding pathway, PDZ domain, polypeptide, transition state.
Abbreviations used: MD, molecular dynamics; TS, transition state.
[1]We are pleased to dedicate this paper to our friend Professor Michael T. Wilson (University of Essex, U.K.) on the occasion of the symposium in his honour.
[2]To whom correspondence should be addressed (email maurizio.brunori@uniroma1.it).

The surprising simplicity of protein folding kinetics

When the unfolded state folds into the native conformation, very many non-covalent bonds are formed, the solvent is excluded from the core of the protein and conformational entropy drops as the folded state is approached. Surprisingly, however, when a small globular protein is challenged with increasing concentrations of denaturant, very often the native structure melts following a simple two-state reaction, such that only the native and denatured states are populated [4]. This all-or-none behaviour implies that the interactions stabilizing the native conformation are

Biochem. Soc. Trans. (2012) 40, 429–432; doi:10.1042/BST20110683

energetically coupled and tend to break (and form) co-operatively. The experimental signature of co-operativity in protein folding lies in the sigmoidal reversible-equilibrium transition and the single exponential time course when both the folding and unfolding reactions are followed in a transient experiment [5].

The observed rate constants for small single domain proteins can span several orders of magnitude, ranging from 0.2 s^{-1} in the case of acyl-phosphatase [6] to up to 10^5 s^{-1} in the case of the engrailed homeodomain [7]. As folding occurs more slowly than a diffusion-limited process (the pre-exponential factor being in the order of 10^6–10^8 s^{-1} [8,9]), it may be concluded that it is a barrier-limited process. Thus a two-dimensional free-energy diagram for a two-state folder consistent with kinetic folding and unfolding experiments has a single barrier for the major TS (transition state). Usually, the reaction co-ordinate of such an energy diagram is the Tanford β_T value, which reflects the non-accessible surface area of the TS relative to the denatured and native states (defined respectively as 0 and 1). β_T can be calculated from the dependence of the folding and unfolding rate constants on denaturant concentration (the so-called Chevron plot). For a two-state folder, crucial information about the energetic and structural properties of the TS can be inferred by a combination of transient kinetics, site-directed mutagenesis and computer simulations. Reliable structural information on the folding TS (via the so-called Φ-value analysis, as championed by Fersht and co-workers [10]) and its characteristic β_T value is essential to unveil the set of contacts crucial to yield the native state.

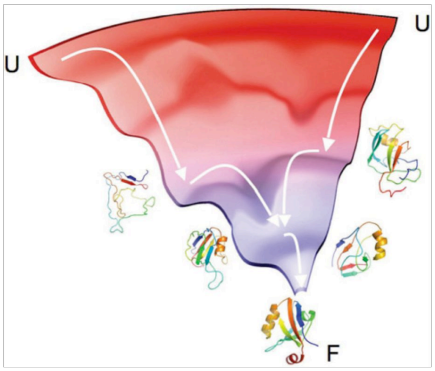
We briefly outline below some experiments relevant to a few basic questions, including the following. Is the denatured state populated under 'native' conditions involved in dictating the final fold? Is it possible to reveal the presence of multiple nuclei at the early stages of folding? Do the structural nuclei of partially folded species along the pathway relate to the structure of the native state?

Plasticity and convergence of folding pathways

Implicit in the view of the energy landscape perspective is that folding occurs via a statistical selection between alternative parallel folding pathways (Figure 1). Experimentally, however, it is not so easy to document alternative folding routes. In fact, over-and-above cases such as lysozyme [11] and cytochrome c_{551} [12] where evidence for more than one path was inferred by double-jump mixing experiments, parallel folding pathways may be hidden in the reaction kinetics since often only one relaxation is experimentally accessible. In these cases, the presence of parallel routes may emerge from a complex analysis of the dependence of the (un)folding rate constant on solvent composition [13]. Quite recently, the ribosomal protein S6 and the PDZ domain have been subjected to extensive protein-engineering experiments that, as discussed below, suggest the presence of alternative nuclei dictating the folding pathway, an observation fully consistent with the energy landscape perspective.

Figure 1 | Schematic representation of a folding funnel

The free energy drives the polypeptide towards the folded state, the most stable conformation, and the conformational entropy dramatically decreases as the native state is approached. Implicit in this view is that folding may occur via alternative parallel pathways. This concept is graphically depicted using the structures of the early and late folding TSs for different PDZ domains, as obtained in [19].

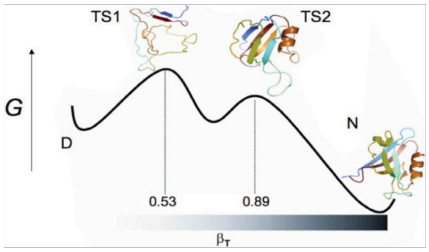


Wild-type S6 is a small globular single-domain protein that, disregarding chain connectivity, is topologically consistent with two specular halves. The folding pathway, characterized by protein engineering and fast kinetics, suggests that the native state is achieved via a diffused TS that encompasses most of the structure; however, a nucleation centre with a detectable content of native-like structure could be identified at the interface between helix1 and the N-terminal strand β1 [14–16]. By systematically altering the sequence connectivity between the six secondary structure elements synthesizing different circularly permuted variants, Oliveberg and co-workers observed that S6 displays a minimal folding nucleus composed of a 'two-strands-one-helix' motif [14]. The exact components of this motif, however, depend on the linear organization of the secondary structure elements, suggesting that circular permutation tunes the balance between competing nuclei involving in all cases the same helix but docking on different strands. This was taken as evidence for pathway malleability in protein folding, with more than one nucleation-competent structural motif within the protein domain. Thus changes in sequence connectivity may alter the relative stability of the folding nuclei and bias the dominant pathway towards alternative routes.

The PDZ domains are a family of proteins that has attracted a lot of attention because of their crucial role in cellular

Figure 2 | Folding free-energy diagram of PDZ2

Representative structures of TS1 and TS2 are reported together with the native state (PDB code 1GM1). Structures were calculated using a combination of Φ-value analysis and MD simulations, as described in [21]. Figure adapted from [21] with permission.



protein–protein interactions and selectivity of binding [17]. Given the ever-growing evidence for the interconnection between binding and folding, PDZ domains have been extensively investigated by kinetics and protein engineering. Curiously the PDZ domains from bacteria and plants may be considered as natural circularly permuted variants of their metazoan counterparts [18]. Hence, they are attractive objects to investigate the relationships between sequence connectivity and protein topology and, for the first time, offer an opportunity to compare engineered and naturally evolved sequences.

The folding of all the PDZ domains studied so far (six canonical and two circular permutants) involves the presence of an intermediate and two sequential TSs (Figure 2) [19–22]. The intermediate may be either high energy as detected for canonical PDZ domains or populated during folding as demonstrated in circularly permuted variants [23–25]. It was particularly interesting to compare the folding pathway of the second PDZ domain from the PTP (protein tyrosine phosphatase) BL (PDZ2) to its engineered circularly permuted variant (cpPDZ2). These two proteins display the same overall structure and sequence composition, but different chain connectivity. Analysis of the structural features of the early and late events in the folding of PDZ2 and cpPDZ2 reveals that while the late stages in the pathway are essentially unaffected by circular permutation, early stages are structurally different, with stabilization of alternative nuclei [24,25]. Thus, in analogy to the S6 proteins, the PDZ domain family display alternative folding pathways involving different nuclei that can be selectively stabilized via loop-entropy perturbations, such as circular permutation. Over-stabilization of a nucleus may lead to frustration of the folding landscape, involving the segregation into local minima that compete for producing the native state. An example of such a scenario is represented by the D1pPDZ domain, a naturally occurring circularly permuted variant that displays an off-pathway kinetic trap characterized by a misfolded N-terminal hairpin incorrectly docked on an otherwise native-like structure [26].

The folding funnel model predicts that the late stages of protein folding are more robust to sequence variations, while the early stages are more malleable [3]. This prediction is difficult to address experimentally, since it requires structural information on the early and late stages of folding, as well as capturing the presence of parallel folding pathways. Once again the PDZ domains proved to be ideal candidates to test such a prediction. Different PDZ domains, such as PDZ2 and PDZ3 (30% sequence identity) both fold via a complex mechanism involving the presence of at least one high-energy folding intermediate, as shown in Figure 2. Early and late events can be addressed experimentally, by probing the structure of the denatured-like TS1 and the native-like TS2 through a combination of Φ-value analysis and restrained MD (molecular dynamics) simulations, to obtain atomic-level structures. It was found that the late TSs of PDZ2 and PDZ3 are structurally similar (TS2 being nearly super-imposable in both cases), while the early ones are quite different [19]. Thus in the case of TS1, limited but detectable native-like structural features could be identified in two different regions, i.e. the $\beta1$–$\beta4$–$\beta6$ interaction (which is present in both PDZ2 and PDZ3) and the $\beta2$–$\beta3$ hairpin (which is dominant in PDZ3). Remarkably, the alternative nucleus involving $\beta2$–$\beta3$ could also be detected in cpPDZ2 [24], suggesting that folding may be rerouted through the alternative pathway via either sequence composition or circular permutation.

Overall, these results suggest that the bias of the native structure is weak at early stages of the folding reaction, allowing for alternative early folding events eventually leading to an intermediate, whereas it is very pronounced in the late stages when the native topology essentially dictates the folding mechanism. In summary, the influence and control of native topology on the structure of intermediate and TSs is loose and permissive at the early stages of folding, but becomes more stringent and effective as the system is approaching the native state, reflecting the funnelling nature of the energy landscape.

Funding

The study was partially supported by the Ministero dell'Istruzione dell'Università e della Ricerca of Italy [grant number RBRN07BMCT_007 to M.B.].

References

1 Anfinsen, C.B., Haver, E., Sela, M. and White, F.H.J. (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. Proc. Natl. Acad. Sci. U.S.A. **47**, 1309–1314
2 Levinthal, C. (1968) Are there pathways for protein folding? J. Chem. Phys. **65**, 44–45
3 Bryngelson, J.D., Onuchic, J.N., Socci, N.D. and Wolynes, P.G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins **21**, 167–195

4 Jackson, S.E. and Fersht, A.R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry **30**, 10428–10435
5 Jackson, S.E. (1998) How do small single-domain proteins fold? Folding Des. **3**, R81–R91
6 van Nuland, N.A., Chiti, F., Taddei, N., Raugei, G., Ramponi, G. and Dobson, C.M. (1998) Slow folding of muscle acylphosphatase in the absence of intermediates. J. Mol. Biol. **283**, 883–891
7 Gianni, S., Guydosh, N.R., Khan, F., Caldas, T.D., Mayor, U., White, G.W., DeMarco, M.L., Daggett, V. and Fersht, A.R. (2003) Unifying features in protein-folding mechanisms. Proc. Natl. Acad. Sci. U.S.A. **100**, 13286–13291
8 Fierz, B., Satzger, H., Root, C., Gilch, P., Zinth, W. and Kiefhaber, T. (2007) Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. Proc. Natl. Acad. Sci. U.S.A. **104**, 2163–2168
9 Kubelka, J., Hofrichter, J. and Eaton, W.A. (2004) The protein folding 'speed limit'. Curr. Opin. Struct. Biol. **14**, 76–88
10 Matouschek, A., Kellis, Jr, J.T., Serrano, L., Bycroft, M. and Fersht, A.R. (1990) Transient folding intermediates characterized by protein engineering. Nature **346**, 440–445
11 Kiefhaber, T. (1995) Kinetic traps in lysozyme folding. Proc. Natl. Acad. Sci. U.S.A. **92**, 9029–9033
12 Travaglini-Allocatelli, C., Gianni, S. and Brunori, M. (2004) A common folding mechanism in the cytochrome *c* family. Trends Biochem. Sci. **29**, 535–541
13 Wright, C.F., Lindorff-Larsen, K., Randles, L.G. and Clarke, J. (2003) Parallel protein-unfolding pathways revealed and mapped. Nat. Struct. Biol. **10**, 658–662
14 Lindberg, M., Tangrot, J. and Oliveberg, M. (2002) Complete change of the protein folding transition state upon circular permutation Nat. Struct. Biol. **9**, 818–822
15 Hubner, I.A., Lindberg, M., Haglund, E., Oliveberg, M. and Shakhnovich, E.I. (2006) Common motifs and topological effects in the protein folding transition state. J. Mol. Biol. **359**, 1075–1085
16 Lindberg, M.O. and Oliveberg, M. (2007) Malleability of protein folding pathways: a simple reason for complex behaviour. Curr. Opin. Struct. Biol. **17**, 21–29
17 Jemth, P. and Gianni, S. (2007) PDZ domains: folding and binding. Biochemistry **46**, 8701–8708
18 Liao, D.I., Qian, J., Chisholm, D.A., Jordan, D.B. and Diner, B.A. (2000) Crystal structures of the photosystem II D1 C-terminal processing protease Nat. Struct. Biol. **7**, 749–753
19 Calosci, N., Chi, C.N., Richter, B., Camilloni, C., Engstrom, Å., Eklund, L., Travaglini-Allocatelli, C., Gianni, S., Vendruscolo, M. and Jemth, P. (2008) Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. Proc. Natl. Acad. Sci. U.S.A. **105**, 19241–19246
20 Gianni, S., Calosci, N., Aelen, J.M., Vuister, G.W., Brunori, M. and Travaglini-Allocatelli, C. (2005) Kinetic folding mechanism of PDZ2 from PTP-BL. Protein Eng. Des. Sel. **18**, 389–395
21 Gianni, S., Geierhaas, C.D., Calosci, N., Jemth, P., Vuister, G.W., Travaglini-Allocatelli, C., Vendruscolo, M. and Brunori, M. (2007) A PDZ domain recapitulates a unifying mechanism for protein folding. Proc. Natl. Acad. Sci. U.S.A. **104**, 128–133
22 Haq, S.R., Jürgens, M.C., Chi, C.N., Koh, C.S., Elfström, L., Selmer, M., Gianni, S. and Jemth, P. (2010) The plastic energy landscape of protein folding: a triangular folding mechanism with an equilibrium intermediate for a small protein domain. J. Biol. Chem. **285**, 18051–18059
23 Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M. and Gianni, S. (2008) Folding and misfolding in a naturally occurring circularly permuted PDZ domain. J. Biol. Chem. **283**, 8954–8960
24 Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M. and Gianni, S. (2009) Engineered symmetric connectivity of secondary structure elements highlights malleability of protein folding pathways. J. Am. Chem. Soc. **131**, 11727–11733
25 Ivarsson, Y., Travaglini-Allocatelli, C., Morea, V., Brunori, M. and Gianni, S. (2008) The folding pathway of an engineered circularly permuted PDZ domain. Protein Eng. Des. Sel. **21**, 155–160
26 Gianni, S., Ivarsson, Y., De Simone, A., Travaglini-Allocatelli, C., Brunori, M. and Vendruscolo, M. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. Nat. Struct. Mol. Biol. **17**, 1431–1437

Received 21 July 2011
doi:10.1042/BST20110683

Folding pathways of proteins with increasing degree of sequence identities but different structure and function

Rajanish Giri[a,1], Angela Morrone[a,1], Carlo Travaglini-Allocatelli[a], Per Jemth[b], Maurizio Brunori[a,2], and Stefano Gianni[a,2]

[a]Istituto Pasteur-Fondazione Cenci Bolognetti, Istituto di Biologia e Patologia Molecolari del Consiglio Nazionale delle Ricerche, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli," Università di Roma "La Sapienza," Piazzale A. Moro 5, 00185 Rome, Italy; and [b]Department of Medical Biochemistry and Microbiology, Uppsala University, Biomedical Center Box 582, SE-75123 Uppsala, Sweden

Edited by William A. Eaton, National Institutes of Health-NIDDK, Bethesda, MD, and approved April 23, 2012 (received for review February 22, 2012)

Much experimental work has been devoted in comparing the folding behavior of proteins sharing the same fold but different sequence. The recent design of proteins displaying very high sequence identities but different 3D structure allows the unique opportunity to address the protein-folding problem from a complementary perspective. Here we explored by Φ-value analysis the pathways of folding of three different heteromorphic pairs, displaying increasingly high-sequence identity (namely, 30%, 77%, and 88%), but different structures called G_A (a 3-α helix fold) and G_B (an α/β fold). The analysis, based on 132 site-directed mutants, is fully consistent with the idea that protein topology is committed very early along the pathway of folding. Furthermore, data reveals that when folding approaches a perfect two-state scenario, as in the case of the G_A domains, the structural features of the transition state appear very robust to changes in sequence composition. On the other hand, when folding is more complex and multistate, as for the G_Bs, there are alternative nuclei or accessible pathways that can be alternatively stabilized by altering the primary structure. The implications of our results in the light of previous work on the folding of different members belonging to the same protein family are discussed.

kinetics | protein engineering | protein folding | protein G

The ultimate goal of a biophysical study is to extract general rules from the analysis of simple systems, a task that is particularly challenging in the case of protein folding. In fact, when considering different globular proteins, complexity stems by and large from the difference in sequence but also the multiplicity of structures of the native and denatured states. A suitable strategy to tackle the problem is to study proteins that differ in sequence but share the same overall fold, i.e., members of the same protein family (1–8). Over the past two decades, this approach allowed drawing some general conclusions about the correlation between 3D structure and sequence composition. In particular, it was reported that the mechanism of folding is, generally, conserved for members of the same protein family (9, 10) supporting the idea that native topology is often a main factor in controlling folding pathway and speed (11).

A sophisticated protein engineering approach allowed Bryan and co-workers to obtain pairs of proteins with an increasing degree of sequence identity (up to the extraordinary value of 95%), but different 3D structure and function (12, 13). The primary structure of two domains from the streptococcal protein G, sharing 16% sequence identity, was subjected to extensive site-directed mutagenesis cycles, leading to the synthesis of pairs of variants with an increasing level of sequence identity (30%, 77%, 88%, and 95%, respectively) (12, 13). The two wild-type protein domains are called G_A, displaying a three-helix bundle fold, and G_B, displaying a $\alpha+\beta$ ubiquitin-like fold. Therefore, the different variants (Fig. 1) were identified as G_A30, G_A77, G_A88, and G_A95 (for the G_A fold), and G_B30, G_B77, G_B88, and G_B95 (for the G_B fold), depending on their relative degree of sequence identities. These five pairs represent a paradigmatic experimental model system to address the folding problem from an original perspective; how can similar sequences lead to very different folds?

We have recently analyzed the folding mechanisms of G_A88 and G_B88 at a variety of different conditions by experiments and simulations (14). We observed that despite the high level of identity of the primary structures of these two proteins (49 out of 56 residues), their folding pathways appear to diverge as early as in the denatured state, which in G_B88 but not in G_A88 displays a detectable residual structure. This surprising finding prompted us to carry out a systematic analysis of each of the heteromorphic variants designed by Bryan and co-workers (12). The natural wild-type G_A domain contains no intrinsic fluorescent probe (i.e., no Trp) and differs by only three amino acids from the lower tier G_A30. Furthermore, we could not carry out a complete folding characterization of the variants G_A95 and G_B95, which differ only in three positions (15), because of their low thermodynamic stability. Therefore, we have undertaken a detailed analysis of the folding pathways of the three remaining heteromorphic pairs, which were subjected to an experimental investigation using the Φ-value analysis (16). This procedure, introduced and validated by Fersht et al. (ref. 17 and references therein), infers structural information on folding transition state(s) by comparing the kinetics and thermodynamics of folding of a given protein with those of a series of conservative mutants; in this work, 132 mutants were fully characterized.

The results show that when folding approaches a perfect two-state scenario, as in the case of G_A, the structural features of the transition state of G_A30, G_A77, and G_A88 appear very robust to changes in sequence composition and are therefore similar. On the other hand, when folding is more complex, as in the case of the three-state folder G_B, there are multiple and alternative nuclei or accessible pathways that can be selectively stabilized by altering the primary structure. The implications of our results in the light of previous work on the folding of members of the same protein families are discussed and highlight the crucial physicochemical features which bias the early commitment to the $\alpha+\beta$ fold of the G_B family.

Results

To test how sequence composition versus topology dictates the folding of proteins, we performed an extensive Φ-value analysis

Author contributions: M.B. and S.G. designed research; R.G. and A.M. performed research; R.G., A.M., C.T.-A., P.J., M.B., and S.G. analyzed data; and M.B. and S.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]R.G. and A.M. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: maurizio.brunori@uniroma1.it or stefano.gianni@uniroma1.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1201794109/-/DCSupplemental.

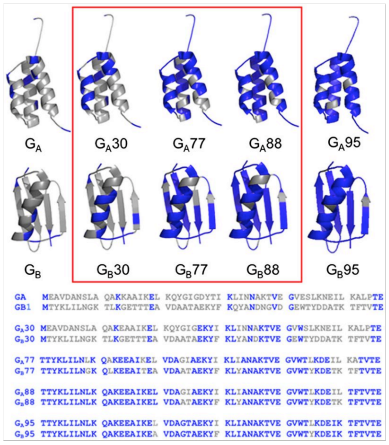


Fig. 1. Structures and sequence alignments of the different G_A and G_B variants. All engineered G_A and G_B proteins display structure and function similar to their respective natural wild-type domains GA and GB1. For each protein, amino acid identities are shown in blue and nonidentities in gray. A complete folding characterization of the variants shown inside the red rectangle was performed.

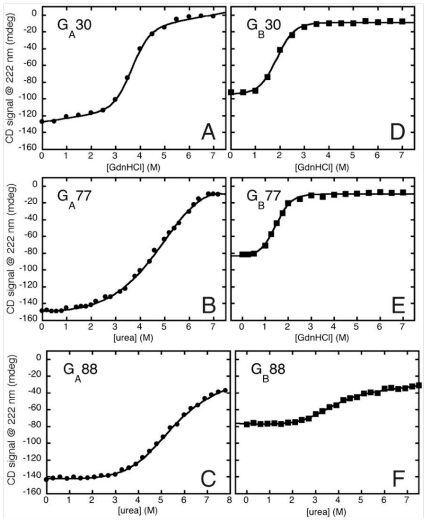


Fig. 2. Equilibrium denaturation of the G_A (*A–C*) and G_B (*D–F*) proteins monitored by CD in 50 mM sodium phosphate buffer at pH 7.2. As described in the text, because of the different thermodynamic stabilities of the proteins, the equilibrium denaturations were performed at different experimental conditions: the unfolding experiment of G_A30 was performed at 25 °C using GdnHCl; G_A77 and G_A88 experiments were performed at 10 °C using urea; the unfolding denaturations of G_B30 and G_B77 were carried out at 25 °C using GdnHCl; and G_B88 experiment was performed at 25 °C using urea and 0.4 M sulfate. For all proteins, the pH was held at the constant value of 7.2 in the presence of 50 mM sodium phosphate.

on six different proteins (Fig. 1), three variants of the G_A domain of staphylococcal protein G (i.e., G_A30, G_A77, and G_A88), and three variants of the G_B domain (i.e., G_B30, G_B77, and G_B88). A total of 132 mutants were produced, purified, and characterized by equilibrium and kinetic folding experiments.

Performing a complete Φ-value analysis demands a careful selection of the experimental conditions. In fact, the protein of interest must be stable enough to allow accurate determination of the folding kinetics for its destabilized mutants, but it must not be too stable, otherwise its unfolding kinetics may be difficult to evaluate with the necessary accuracy. As detailed below, since the three pairs of proteins characterized in this work and their site-directed variants displayed widely different thermodynamic stabilities as well as different sensitivities to changes in ionic strength, we optimized the experimental conditions for each protein with regard to temperature and type of denaturant (i.e., urea or GdnHCl). In addition, in an effort to test the robustness of the folding pathway, a limited set of Φ values was measured for each protein at more than one experimental condition, while the full set of Φ values was obtained for each system at pH 7.2 (see *SI Text*).

***G_A* Proteins.** Urea induced equilibrium transitions of G_A30, G_A77, and G_A88 were measured at 25 °C and pH 7.2 in 50 mM sodium phosphate buffer. A typical equilibrium denaturation profile for each protein is reported in Fig. 2 *A–C*). While in the case of G_A77 and G_A88 we could observe monotonic sigmoidal transitions, G_A30 was found to be too stable and was not fully denatured even at very high urea concentrations; therefore, in the case of G_A30 we performed (un)folding experiments using GdnHCl. Because of the ionic nature of GdnHCl, we could not perform (un)folding experiments with this denaturant on G_A77 and G_A88, since their stabilities display a pronounced dependence on ionic strength, as shown by folding experiments in the presence of sodium chloride.

We extensively studied the folding and unfolding kinetics of G_A30, G_A77 and G_A88 by stopped-flow and temperature jump (T-jump) experiments. In the case of G_A77 and G_A88, it was not possible to measure reliable folding and unfolding rate constants at 25 °C over a wide range of denaturant concentration because reactions were too fast for our stopped-flow apparatus; thus, kinetic folding data for the two proteins were recorded at 10 °C. In all cases, the folding and unfolding time courses were fitted satisfactorily to a single exponential decay at any final denaturant concentration. Semilogarithmic plots of the observed folding/unfolding rate constants of G_A30, G_A77, and G_A88 versus denaturant concentration (i.e., chevron plots) are presented in Fig. 3 *A–C*. All proteins displayed a V-shaped chevron plot, a hallmark of two-state folding. Two-state folding was further confirmed by the excellent agreement between the thermodynamic parameters obtained by equilibrium and kinetic data. These data parallel earlier studies on G_A88 (14) and confirm that this protein system folds via a two-state folding pathway with an unstructured denatured state.

We addressed the structural features of the transition state for folding of the G_A proteins by Φ-value analysis. A total of 50 mutants was produced: 14 for G_A30, 17 for G_A77, and 19 for G_A88. Three mutants expressed poorly or were too unstable to be included in the analysis. The remaining 47 were subjected to equilibrium and kinetic folding experiments (Figs. S1–S3). In some cases, the folding and unfolding rate constants were too fast for the stopped-flow methodology and were determined using a capacitor-discharge T-jump apparatus. Fitted parameters are listed in Table S1.

Following a generally accepted convention (16, 18–20), the experimentally determined Φ-values were grouped in three different classes and mapped on the native structure of the G_A protein (Fig. 4): small values (Φ < 0.3; red), intermediate values (0.3 < Φ < 0.7; magenta), and large values (Φ > 0.7; blue). Inspection

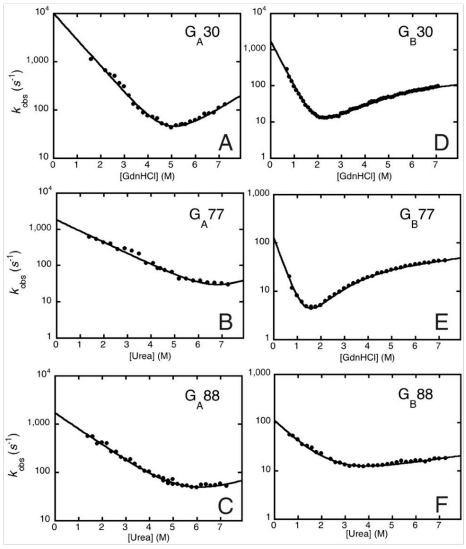


Fig. 3. Chevron plots of G_A and G_B variants. (*A–C*) panels: semilogarithmic plots of the observed rate constant for folding and unfolding of G_A30, G_A77, and G_A88 versus denaturant concentration measured at pH 7.2 and 10 °C. (*D–F*) panels: Semilogarithmic plots of the observed rate constant for folding and unfolding of G_B30, G_B77, and G_B88 versus denaturant concentration measured at pH 7.2 and 25 °C. The lines are the best fit to a two- or a three-state model.

of Fig. 4 clearly reveals that the folding nucleus is by and large conserved in all the G_A proteins, with the highest Φ values clustered at the interface between helix 1 and helix 2, suggesting that, in the case of such a two-state system, the dominant folding mechanism is very robust to perturbations of the primary structure.

G_B Proteins. The folding pathway of the G_B proteins is inherently more complex than that of the G_A partners. In fact, we have previously shown (21) that the wild-type G_B protein, commonly referred as GB1, a widely used model system for protein-folding studies, is characterized by the presence of an on-pathway intermediate, as indicated by a curvature in the unfolding arm of its chevron plot. Such a curvature, detected only at high concentrations of GdnHCl, allows addressing experimentally both the early and late folding events. Consequently, we resorted to study the folding of G_B30, G_B77, and G_B88 at 25 °C and pH 7.2 in 50 mM sodium phosphate buffer, using GdnHCl as a denaturing agent. Unfortunately, however, in the case of G_B88, many of its site-directed variants were poorly soluble at moderate concentrations of GdnHCl, and we could not obtain reliable folding data on this system with this denaturant. Furthermore, the low stability of this variant did not allow us to perform a complete Φ-value analysis in the absence of the stabilizing agent sodium sulfate. All folding studies of G_B88 were therefore carried out using urea as denaturant and in the presence of 0.4 M sodium sulfate.

The equilibrium unfolding transitions of G_B30, G_B77, and G_B88 are reported in Fig. 2 *D–F*. In all cases, we observed a monotonic sigmoidal transition, suggesting the absence of equilibrium intermediates.

The chevron plots for G_B30, G_B77, and G_B88 as a function of denaturant concentration, at 25 °C and pH 7.2, are reported in Fig. 3 *D–F*. It is evident that, in analogy to what previously

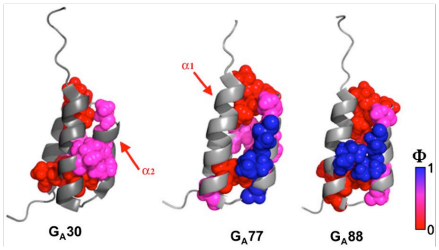


Fig. 4. Structural distribution of the measured Φ-values on the native structures of G_A30, G_A77, and G_A88. The experimentally determined Φ values were divided into three categories and reported on the structure of G_A variants using the following color code: red, 0 < Φ < 0.30; magenta, 0.30 < Φ < 0.70; blue, 0.70 < Φ < 1. A conserved nucleus between the helices α 1 and α 2 (indicated by the red arrows) is clearly evident in all the G_A variants.

observed for GB1, both G_B30 and G_B77 display a pronounced curvature in their unfolding arms, as expected for a three-state folding mechanism (22, 23). In the case of G_B88, while we could not detect such a curvature when performing the experiments in urea, data recorded in the presence of high concentrations of GdnHCl were similar to those of G_B30 and G_B77 (Fig. S4). This observation suggests that all the G_B variants appear to conform to a three state-folding mechanism.

We addressed the structural features of the early and late transition states for folding of the G_B proteins by Φ-value analysis. We produced a total of 82 mutants: 24 for G_B30, 27 for G_B77, and 31 for G_B88. 16 variants expressed poorly or were too unstable to be included in the analysis. The remaining 66 were subjected to equilibrium and kinetic folding experiments (Figs. S5–S7). In the case of G_B30 and G_B77, the data for each variant was globally fitted to a three-state equation with *m*-values assumed to be the same as those of the wild-type proteins. All data was in agreement with a three-state mechanism involving the presence of two folding transition states, a more denatured-like *TS1* ($\beta = 0.77$ and $\beta = 0.69$ for G_B30 and G_B77, respectively) and a native-like *TS2* ($\beta = 0.97$ for both the proteins). On the other hand, in the case of G_B88, because of the low solubility of its variants in the presence of GdnHCl, we could perform the experiments only in the presence of urea, obtaining structural information only on the early transition state *TS1* ($\beta = 0.81$).

A graphic depiction of the structural distribution of the measured Φ-value for *TS1* and *TS2* of G_B30 and G_B77 and of *TS1* of G_B88 is reported in Fig. 5. It is interesting to note that the distribution of the measured Φ-values for the first transition state (Φ_{TS1}), plotted on the native structure of G_B, show considerable differences among G_B30, G_B77, and G_B88. In fact, looking at G_B30 and G_B88, a shift of the medium-high Φ-values from the first β-hairpin to the second, with G_B77 displaying an intermediate behavior (Fig. 5), may be appreciated. This trend indicates that alternative folding nuclei, located at the hairpins between either β1-β2 or β3-β4, drive the folding to the G_B-topology. These nuclei may be selectively stabilized depending on amino acid composition. Remarkably, in the native-like transition state *TS2*, which we could infer only for G_B30 and G_B77, both nuclei appear in the process of being folded, and the two transition states display a similar overall structure, indicating that the alternative folding pathways converge as the native state is approached. In order to compare the Φ-values obtained under different conditions, kinetic experiments on some mutants were performed with both guanidine and urea. The calculated Φ-values were approximately the same; the folding mechanism of the G_B proteins is therefore not affected by the nature of the denaturant agent used to obtain the data [Fig. S8 shows the structural distribution

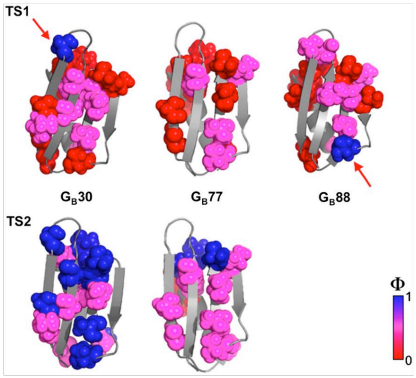


Fig. 5. Structural distribution of the measured Φ-values for *TS1* (G_B30, G_B77, and G_B88) and for *TS2* (G_B30 and G_B77). The experimentally determined Φ values were divided into three categories and reported on the native structure of G_B variants using the same color code as that used for the G_A proteins. In the case of G_B30 and G_B88, it is possible to observe a shift of the medium-high Φ-values from the first β-hairpin to the second (indicated by the red arrows), with G_B77 displaying an intermediate behavior.

of the measured Φ-values for *TS1* of G_B77 obtained in presence of GdnHCl (A) and urea (B)].

Discussion

The holy grail in protein-folding studies is to unveil the correlation between sequence information and reaction mechanisms. A classical approach to address this question has been to study proteins that differ in amino acid sequence but share the same fold (1–8). On the other hand, the design and production of proteins sharing high sequence identity, yet displaying a different structure and function, allows for the first time to approach the folding problem from a complementary perspective (12, 15, 22, 23). In this work, we have characterized the complete folding pathway of heteromorphic proteins originating from two domains of streptococcal protein G being either fully α-helical (called G_A) or largely β (called G_B). An extensive Φ-value analysis of G_A30, G_A77, G_A88, and of G_B30, G_B77 and G_B88 generates a "matrix" of proteins, which share the same topology (the two rows of Fig. 1) and display increasingly high sequence identities while keeping a different structure.

When does a protein commit to its native topology in its folding pathway? In a preliminary study, we compared the folding and unfolding kinetics of G_A88 and G_B88 at a variety of different experimental conditions (14). We observed that a detectable residual structure is present in the denatured state of G_B88, while the denatured state of G_A88 is essentially unstructured. In the light of these observations, we hinted that the folding pathways of these two proteins diverge as early as in the denatured state. In the present work, by systematically describing the structural features of the folding transition states of G_A30, G_A77, G_A88, as well as of G_B30, G_B77, and G_B88, we have taken our comparative study on heteromorphic pairs to the next level of complexity addressing by Φ-value analysis the structure of the transition states. Remarkably, the structural distribution of the measured Φ-values in all these variants is reminiscent of what has been previously observed in many single domain proteins, with a weak nucleus displaying moderately high values of Φ (>0.7), and simultaneous formation of extensive native-like structure, which gradually tapers off from the nucleus to other regions of the protein (24, 25). Accordingly, the transition states of all the proteins considered here appear to resemble a distorted version of the corresponding native states, with some polarization of structure at the N- or C-terminal β-hairpins in the case of G_B30 and G_B88 respectively. Furthermore, additional support for the presence of residual structure in the denatured state of G_B88 is represented by the presence of three nonstandard values of Φ (i.e., L20A, V21A, and Y33F; Table S2). Remarkably, these positions are located at the N- and C-terminal regions of the central helix of the G_B fold, which was observed by molecular dynamics simulation to be partially formed in the denatured state of G_B88 (14). Overall, despite the very high level of sequence identities, when comparing the folding of each heteromorphic pair, no common intermediate was detected, suggesting that they fold via completely independent paths. All these findings converge in indicating that proteins commit very early to their topology, and the structural features leading to their native state are, most likely, already imprinted in their denatured states.

A crude but reliable method to compare the folding pathways of these different proteins is to analyze the structural distribution of measured Φ-values. From this perspective, data indicates that, while in the case of G_A the mechanism appears rather robust to divergence in primary structure, in the case of G_B, folding is more malleable. In fact, when mapped onto the corresponding native structures (Fig. 4), the Φ-values reveal a conserved transition state among all the three G_A variants (G_A30, G_A77, G_A88); on the other hand, G_B30 and G_B88 clearly display a shift of the medium-high Φ-values from the first β-hairpin to the second, with G_B77 displaying an intermediate outlook (Fig. 5). This finding may be seen in the light of the proposal that the number of accessible pathways for folding is determined by the different nucleation motifs contained within a given native topology (26). For example, the structure of ribosomal protein S6 seems to be composed of two different nucleation patterns acting as independent cooperative units, each of which constitutes a separate entry to parallel folding trajectories (27). Accordingly, it may be suggested that in the case of G_B, the symmetrical organization of its 3D topology implies the presence of multiple nucleation motifs that permit alternative folding pathways.

For our purposes, it would be revealing to understand which structural determinants preclude the sequence of the G_A proteins to adopt the structure of G_B and vice versa. Recent molecular dynamics simulations on G_A88 and G_B88 suggested the long and stable helix in the central region of the sequence of G_B88 preventing the polypeptide chain to form the loop connecting helix 1 and helix 2 in G_A88 and thus to fold into a fully helical structure (14). This finding is further corroborated by the high helical propensity of the only α-helix of all the G_B proteins as predicted by AGADIR (28), when compared to the G_A counterparts. By following this view, it is interesting to note that structure selection in the G_A proteins seems to be initiated by the formation of the critical contacts between helix 1 and helix 2 (Fig. 4). In the case of the G_B proteins, on the other hand, the Φ-value analysis reported in this work, together with the previously published molecular dynamics simulations, suggests a scenario whereby folding is guided by the alternative docking of the N- or C-terminal hairpins on the central long helix, that might be (partially) preformed in the denatured state.

Although the protein-folding reaction involves the formation and breakage of a myriad of contacts, a typical feature of small single domain proteins is the ability to fold cooperatively (29–31). Many weak noncovalent bonds form simultaneously and, very often, only the fully native and fully denatured states may be experimentally detected. Yet, because cooperativity is never extreme, not all residues are equally important for folding, and one or more regions of the protein tend to display a selective propensity to fold independently (32, 33). These subdomains initiate the folding reaction and act as folding nuclei. Accordingly, when cooperativity decreases, the energetic coupling for folding of different regions of the same protein is also decreased, i.e., different regions of the protein may fold independently and folding

may become modular. In this perspective, it is of interest to compare the different robustness of the folding mechanisms for the G_A and G_B proteins. In fact, while G_A, a two-state system, appears to display a unique nucleus, in the case of G_B, where the tendency to populate folding intermediates is documented, alternative nuclei are present and folding is more sensitive to changes in sequence composition. Overall, our data reveals that pathway malleability is determined by the presence of multiple nuclei; the segregation of such nuclei results in stabilization of folding intermediates, whereas "perfect" two-state systems are characterized by a unique diffused nucleus and, therefore, by a robust folding pathway.

Materials and Methods

Site-Directed Mutagenesis and Protein Expression and Purification. G_A and G_B genes were cloned into the vector pG58 (provided by Philip N. Bryan, University of Maryland) which encodes an engineered subtilisin pro-sequence as the N terminus of the fusion protein (34). These genes were used as templates to perform site-directed mutagenesis. All mutants were obtained by using the QuikChange mutagenesis kit (Stratagene), according to the manufacturer's instructions. All of the mutations were confirmed by DNA sequencing.

Proteins were expressed and purified as described previously (14).

Equilibrium Unfolding. Circular dichroism (CD) spectra were recorded between 250 and 200 nm using a Jasco spectropolarimeter (Jasco, Inc., Easton, MD, USA). Urea and guanidine denaturations were followed at 222 nm in a 1 cm path length quartz cuvette (Hellma, Plainview, NY, USA) at 10 °C or 25 °C. Protein concentration was typically 15 μM. The buffer used was 50 mM sodium phosphate pH 7.2.

Stopped-Flow Measurements. Single mixing kinetic folding experiments were carried out on a Pi-star or on an SX-18 stopped-flow instruments (Applied Photophysics). The excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. In all experiments, performed at 25 °C and 10 °C, refolding and unfolding were initiated by an 11-fold dilution of the denatured or the native protein with the appropriate buffer. The buffer used was 50 mM sodium phosphate pH 7.2. Final protein concentrations were typically 1 μM. The observed kinetics were always independent of protein concentration (from 0.5 to 5 μM after mixing protein concentration), as expected from monomolecular reactions without effects due to transient aggregation (35).

T-Jump Fluorescence Spectroscopy. The relaxation kinetics were measured as a function of guanidine or urea by using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (Hi-Tech). Temperature was rapidly changed from 18 °C to 25 °C and from 4 °C to 10 °C with a jump size of 7 °C and 6 °C, respectively. 10 to 20 individual traces were averaged at given denaturant concentrations. Protein concentration was typically 20 μM. The excitation wavelength was 280 nm, and the fluorescence emission was measured using a 320 nm cut-off glass filter.

Data Analysis. ***Equilibrium experiments.*** Data were fitted to a standard two-state denaturation. An equation that takes into account the pre- and post-transition baselines was used to fit the observed unfolding transition (36).

Kinetic experiments. Analysis was performed by nonlinear least squares fitting of single exponential phases using the fitting procedures provided in the Applied Photophysics software. The chevron plots were fitted globally by numerical analysis based on a two- or a three-state model.

The logarithm of each microscopic rate constant was assumed to vary linearly with denaturant concentration. The observed chevron plots were fitted globally with shared *m*-values. The global fit was obtained with Prism software (Graphpad). Φ values were calculated from folding rate constants using standard equations (17).

ACKNOWLEDGMENTS. We thank Prof. Philip N. Bryan for inspiring this work. Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (RBRN07BMCT_007 to M.B) Progetto di Interesse 'Invecchiamento' (S.G.), and from the Swedish Research Council (P.J.).

1. Calosci N, et al. (2008) Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc Natl Acad Sci USA* 105:19241–19246.
2. Chi CN, et al. (2007) A conserved folding mechanism for PDZ domains. *FEBS Lett* 581:1109–1113.
3. Clarke J, Cota E, Fowler SB, Hamill SJ (1999) Folding studies of Ig-like beta-sandwich proteins suggest they share a common folding pathway. *Structure* 7:1145–1153.
4. Friel CT, Capaldi AP, Radford SE (2003) Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: Similarities and differences in the folding of homologous proteins. *J Mol Biol* 36:293–305.
5. Martínez JC, Serrano L (1999) The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat Struct Biol* 6:1010–1016.
6. Riddle DS, et al. (1999) Experiment and theory highlight role of native state topology in SH3 folding. *Nat Struct Biol* 6:1016–1024.
7. Travaglini-Allocatelli C, Gianni S, Brunori M (2004) A common folding mechanism in the cytochrome c family. *Trends Biochem Sci* 29:535–541.
8. Travaglini-Allocatelli C, et al. (2003) Exploring the cytochrome c folding mechanism: Cytochrome c552 from Thermus thermophilus folds through an on-pathway intermediate. *J Biol Chem* 278:41136–41140.
9. Travaglini-Allocatelli C, Ivarsson Y, Jemth P, Gianni S (2009) Folding and stability of globular proteins and implications for function. *Curr Opin Struct Biol* 19:3–7.
10. Zarrine-Afsar A, Larson SM, Davidson AR (2005) The family feud: Do proteins with similar structures fold via the same pathway? *Curr Opin Struct Biol* 15:42–49.
11. Baker D (2000) A surprising simplicity to protein folding. *Nature* 405:39–42.
12. Alexander PA, et al. (2007) The design and characterization of two proteins with 88% sequence identity but different structure and function. *Proc Natl Acad Sci USA* 104:11963–11968.
13. He Y, Chen Y, Alexander P, Bryan PN, Orban J (2008) NMR structures of two designed proteins with high sequence identity but different fold and function. *Proc Natl Acad Sci USA* 105:14412–14417.
14. Morrone A, et al. (2011) The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. *J Biol Chem* 286:3863–3872.
15. Alexander PA, et al. (2009) A minimal sequence code for switching protein structure and function. *Proc Natl Acad Sci USA* 106:21149–21154.
16. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224:771–782.
17. Fersht AR, Sato S (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc Natl Acad Sci USA* 101:7976–7981.
18. Geierhaas CD, Salvatella X, Clarke J, Vendruscolo M (2008) Characterisation of transition state structures for protein folding using 'high,' 'medium,' and 'low' {Phi}-values. *Protein Eng Des Sel* 21:215–222.
19. Gianni S, et al. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat Struct Mol Biol* 17:1431–1437.
20. Ivarsson Y, Travaglini-Allocatelli C, Brunori M, Gianni S (2009) Engineered symmetric connectivity of secondary structure elements highlights malleability of protein folding pathways. *J Am Chem Soc* 131:11727–11733.
21. Morrone A, et al. (2011) Gb1 is not a two-state folder: Identification and characterization of an on-pathway intermediate. *Biophys J* 101:1–8.
22. Parker MJ, Spencer J, Clarke AR (1995) An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J Mol Biol* 253:771–786.
23. Wildegger G, Kiefhaber T (1997) Three-state model for lysozyme folding: Triangular folding mechanism with an energetically trapped intermediate. *J Mol Biol* 270:294–304.
24. Abkevich VI, Gutin AM, Shakhnovich EI (1994) Specific nucleus as the transition state for protein folding: Evidence from the lattice model. *Biochemistry* 33:10026–10036.
25. Itzhaki LS, Otzen DE, Fersht AR (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol* 254:260–288.
26. Lindberg MO, Oliveberg M (2007) Malleability of protein folding pathways: A simple reason for complex behavior. *Curr Opin Struct Biol* 17:21–29.
27. Haglund E, Lindberg MO, Oliveberg M (2008) Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity. *J Biol Chem* 283:27904–27915.
28. Munoz V, Serrano L (1997) Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* 41:495–509.
29. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21:167–195.
30. Fersht AR (2008) From the first protein structures to our current knowledge of protein folding: Delights and scepticisms. *Nat Rev Mol Cell Biol* 9:650–654.
31. Kubelka J, et al. (2006) Sub-microsecond protein folding. *J Mol Biol* 359:546–553.
32. Fersht AR (1995) Optimization of rates of protein folding: The nucleation–condensation mechanism and its implications. *Proc Natl Acad Sci USA* 21:10869–10873.
33. Munoz V, Thompson PA, Hofrichter J, Eaton WA (1997) Folding dynamics and mechanism of beta-hairpin formation. *Nature* 390:196–199.
34. Ruan B, et al. (2004) Engineering subtilisin into a fluoride-triggered processing protease useful for one-step protein purification. *Biochemistry* 43:14539–14546.
35. Silow M, Oliveberg M (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc Natl Acad Sci USA* 94:6084–6086.
36. Santoro MM, Bolen DW (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* 27:8063–8068.

THE PROTEIN SOCIETY

FOR THE RECORD

Reassessing the folding of the KIX domain: Evidence for a two-state mechanism

Angela Morrone, Rajanish Giri, Maurizio Brunori, and Stefano Gianni*

Istituto Pasteur-Fondazione Cenci Bolognetti, Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, Piazzale A. Moro 5, 00185 Rome, Italy

Received 6 July 2012; Accepted 11 September 2012
DOI: 10.1002/pro.2159
Published online 25 September 2012 proteinscience.org

Abstract: The debate about the presence and role of intermediates in the folding of proteins has been a critical issue, especially for fast folders. One of the classical methodologies to identify such metastable species is the "burst-phase analysis," whereby the observed signal amplitude from stopped-flow traces is determined as a function of denaturant concentration. However, a complication may arise when folding is sufficiently fast to jeopardize the reliability of the stopped-flow technique. In this study, we reassessed the folding of the KIX domain from cAMP Response Element-Binding (CREB)-binding protein, which has been proposed to involve the formation of an intermediate that accumulates in the dead time of the stopped flow. By using an in-house-built capillary continuous flow with a 50-μs dead time, we demonstrate that this intermediate is not present; the problem arose because of the instrumental limitation of the standard stopped flow to assess very fast refolding rate constants (e.g., ≥500 s^{-1}).

Keywords: protein folding; kinetics; intermediates; continuous flow

Introduction

One of the most important tasks in studying the folding of a protein is to identify the simplest kinetic scheme describing this complex reaction. Some proteins, such as Chymostrypsyn inhibitor 2 (CI2) or the Lysin Motif (LysM) domain, tend to fold in a highly cooperative manner, such that only the fully denatured and native states are stable enough to be populated.[1–4] Some other proteins fold extremely fast and the characterization of their folding pathways may demand the use of ultrarapid techniques, such as the laser T-jump.[5,6] When the stability of individual elements of secondary structure or local nuclei is more marked,[7] proteins form intermediates, whose structure must be elucidated to describe their role in folding.[8]

A classical signature of folding intermediates is represented by multiphasic kinetics. In some cases, however, intermediates are more elusive and their role must be inferred indirectly.[9] A classical test to pin down a folding intermediate is based on the dependence on denaturant of the folding and unfolding rate constants and on burst-phase analysis of stopped-flow experiments.[10] By using the latter methodology, the denaturant dependence of the apparent fluorescence amplitude coupled to refolding yields a quantitative assessment for an intermediate. In the absence of ultrafast events (lost in the dead time of the stopped flow), the fluorescence amplitude of refolding experiments is expected to depend linearly on denaturant

Grant sponsor: Italian Ministero dell'Istruzione dell'Università e della Ricerca; Grant numbers: RBRN07BMCT_007 (M.B.), PNR-CNR Aging Program 2012-2014 (S.G.)

*Correspondence to: Stefano Gianni, Istituto Pasteur-Fondazione Cenci Bolognetti, Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A., Rossi Fanelli", Sapienza Università di Roma, Piazzale A. Moro 5, 00185 Rome, Italy. E-mail: stefano.gianni@uniroma1.it

concentration;[11] sudden increase or decrease of the fluorescence burst-phase is then interpreted as an experimental evidence for a folding intermediate. In these cases, because the intermediate is obviously populated, the observed folding rate constant is slowed down by a factor equal to $1 + K_{D-I}$ and the dependence of the logarithm of the folding rate constant on denaturant concentration is not linear, the chevron plot displaying the so-called rollover effect.[12] Given that small globular proteins often fold very rapidly,[3] folding rates may easily approach the limits of standard stopped-flow instruments; for example, with a folding rate constant of 400 s^{-1}, more than 50% of the total signal change may be lost in a 2-ms dead-time instrument. In these cases, experience suggests that a burst-phase in the refolding amplitudes and a rollover effect in the chevron plot might be vitiated by the limitation of the machine;[13] and therefore, additional experiments may be necessary to validate the presence of intermediates.

The KIX domain is part of a large multidomain protein, the coactivator CREB-binding protein (CBP), which has recently attracted considerable attention. In fact, KIX is capable of recognizing several intrinsically unstructured proteins, such as the phosphorylated kinase inducible domain (pKID) domain from CREB, the transactivation domain of p53, and the activation domain of the hematopoietic transcription factor c-Myb, which all fold on binding.[14–17] The folding and unfolding reactions of KIX have been studied both by equilibrium and by kinetic experiments.[18,19] The analysis of the amplitude and rate constants obtained by refolding experiments suggested the KIX domain to fold via a low-energy intermediate, which accumulates in the dead time of the stopped flow. Because of the low stability of such an intermediate, it is unclear whether this species is the same as that detectable by equilibrium experiments. Since at low urea concentrations, the observed folding rate constants of KIX approach the experimental limitation of the stopped-flow methodology (ca. 400 s^{-1} at 2*M* urea), the rollover effect/burst-phase analysis of this protein demand additional validation. In this work, we have reassessed the folding kinetics of KIX using an in-house-built continuous-flow apparatus with a dead time of about 50 μs. Data reveal that the rollover effect is absent when refolding was monitored by continuous flow, casting some doubts on the existence of the low-energy burst-phase intermediate, which may result from an intrinsic limitation of the stopped-flow instrument employed.

Results and Discussion

The dependence of the folding and unfolding rate constants of KIX versus denaturant concentration is reported in Figure 1, compared with data previously published.[18] While above 2*M* urea the data obtained in this laboratory agree with those previously published, a clear difference was observed at low denaturant concentrations (i.e. at [urea] < 1.5*M*). Prompted by such a discrepancy, we resorted to complement the standard stopped-flow data with experiments carried out with an ultrarapid in-house-built continuous-flow instrument. Because of the high amount of protein required for continuous-flow experiments, in analogy to previous work carried out on the protein Im7,[20] we carried out these experiments with the His-tagged version of KIX. Importantly, the presence of the tag has no significant effect on both the stability and folding kinetics of KIX; Figures 1 and 2 show a comparison between the His-tagged and non-His-tagged proteins, the latter being the same construct used in Ref. 18. A typical refolding trace of KIX measured at 25°C in the presence of 10 m*M* phosphate and 150 m*M* NaCl pH 7.5 and 0.6*M* urea is reported in Figure 2(A). The complete chevron plot from stopped- and continuous-flow experiments, reported in Figure 2(B), shows that the rollover previously seen at low urea concentrations in not visible when the refolding rate constant is measured with an instrument characterized by a much smaller dead time. Thus, it is likely that the rollover effect reflects the intrinsic limitation of the stopped-flow instrument. Furthermore, an underestimate of the refolding rate constant would correspond to an artifactual burst phase in the amplitude analysis.

In an effort to investigate further the presence/absence of a rollover effect in the folding of KIX, we carried out kinetic experiments also in the presence of a stabilizing salt, namely 0.4*M* sodium sulfate

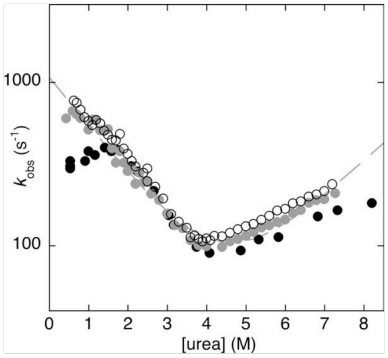


Figure 1. Chevron plot of KIX obtained by stopped flow. Semilogarithmic plot of the observed rate constants for folding and unfolding of KIX versus [urea] at pH 7.5 at 25°C in 10 m*M* NaPi and 150 m*M* NaCl. The chevron plots of KIX protein with His-tag (gray circles) and without His-tag (black open circles) obtained by stopped flow are superposed with that previously published by Horng et al.[18] (black filled circles).

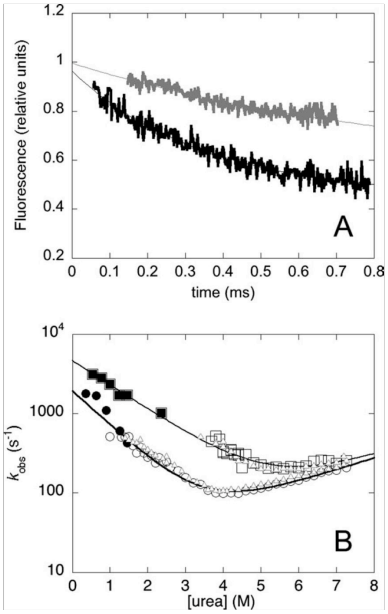


Figure 2. Folding kinetics of KIX from μs to ms by stopped flow and continuous flow. (A) Refolding traces of KIX measured by the continuous-flow apparatus. Both traces were recorded in 10 m*M* NaPi pH 7.5 and 150 m*M* NaCl and 25°C, in the absence (gray, 0.6*M* urea) and in the presence (black, 0.54*M* urea) of 0.4*M* Na_2SO_4. Lines are the best fit to a single exponential decay.
(B) Semilogarithmic plot of the observed rate constants for folding and unfolding of KIX versus [urea] at pH 7.5 in 10 m*M* NaPi and 150 m*M* NaCl, obtained in the presence (squares) and in the absence (circles) of 0.4*M* Na_2SO_4. The data obtained after His-tag cleavage in the presence and absence of 0.4*M* Na_2SO_4 are reported as gray triangles. Lines are the best fit to a two-state chevron plot. Folding rate constants higher than 500 s^{-1} as obtained by stopped flow (Fig. 1) have been omitted from this analysis. The folding rate constants in the absence of denaturant extrapolated to zero urea are 1900 s^{-1} in the absence and 4600 s^{-1} in the presence of 0.4*M* Na_2SO_4.

[Fig. 2(B)]. A typical refolding trace of KIX recorded in presence of 0.4 M sodium sulfate is reported in Figure 2(A). As expected, the salt shifts the denaturation midpoint to higher urea concentrations (from 4*M* in the absence to 5*M* in the presence of 0.4 *M* sodium sulfate) and increases the folding rate constant by a factor of 3. In analogy to what observed in the absence of sodium sulfate, when the data are recorded using the stopped-flow and the continuous-flow instruments, time course is always monoexponential, and there is no indication of refolding rollover effect. Thus, the chevron plot of KIX appears to satisfy a simple two-state folding behavior without accumulation of intermediates.

Concluding Remarks

The KIX domain of CBP has been reported to fold through an early intermediate based on two classical observations: significant burst-phase amplitudes and rollover in the folding limb of the chevron plot.[18] In carrying out a detailed chevron plot analysis by combining stopped-flow and continuous-flow experiments, both in the presence and in the absence of the stabilizing agent sodium sulfate 0.4*M*, we did not detect the previously observed signatures for such an intermediate. Even if there are several examples in the literature of burst-phase intermediates in fast folding protein domains,[10,21–23] this effect must be interpreted with some caution. In fact in some cases, a rollover may arise from protein aggregation,[24] shift of the transition state along the reaction coordinate[25] or inadequate pH control on buffer dilution. In this study, we exemplify that the saturation of the apparent rate constants at low urea may induce apparent rollover effect. In the latter case, ultrafast mixing experiments, classically introduced to monitor the accumulation of short-lived intermediates,[26–30] allow to infer unambiguously the existence of an intermediate, if any. By integrating stopped-flow and continuous-flow mixing experiments, in fact, we obtained for the folding kinetics of KIX a completely linear chevron plot typically diagnostic of a simple two-state model.

Material and Methods

Expression and purification of KIX domain

KIX from Mus musculus CBP (residues 587−672) cloned into pRSET vector (Invitrogen) and resulting in a clone expressing His-tagged KIX, was generously provided by Prof. Per Jemth (Uppsala University, Sweden). The plasmid was used to transform *Escherichia coli* BL-21 (DE3), and protein expression was initiated by inducing 1-L bacterial cultures with 1-m*M* isopropyl-β-D-thiogalactopyranoside at an A600 of ~0.5. The cultures were grown overnight at 25°C, and the bacteria were harvested by centrifugation. KIX was purified by using a nickel(II)-charged chelating Sepharose FF (Amersham Biosciences) column equilibrated with 40 m*M* Tris-HCl and 400 m*M* NaCl, pH 8.5. After washing with the same buffer, the His-tagged Kix was eluted with 250 m*M* imidazole. The sample was then diluted fourfold in 40 m*M* Tris and pH 8.5, and all minor impurities were removed by the purification step on a Q-sepharose column equilibrated with 40 m*M* Tris and pH 8.5. The flow-through containing the protein was collected and concentrated. When necessary, His-tag cleavage was carried out

with thrombin (Sigma Aldrich) using 2.0 U for mg of protein at 4°C for 15 h. KIX was further purified from thrombin using para-aminobenzoic acid (PABA) column (GE Healthcare). The purity of the protein was confirmed by sodium dodecyl sulphate-polyacrilamide gel electrophoresis (SDS-PAGE).

Stopped-flow measurements

Single mixing kinetic folding experiments were carried out on an SX-18 stopped-flow instrument (Applied Photophysics, Leatherhead, UK); the excitation wavelength was 280 nm, and the fluorescence emission was measured using a 320-nm cutoff glass filter. In all experiments, refolding and unfolding were initiated by a 11-fold dilution of the denatured or the native protein with the appropriate buffer. The buffer used was 10 m*M* sodium phosphate, 150 m*M* NaCl, and pH 7.5. Final protein concentration was typically 1 μ*M*. The observed kinetics were always independent of protein concentration (from 0.5 to 5 μ*M* after mixing), as expected from monomolecular reactions without effects due to transient aggregation.[24]

Continuous-flow experiment

Continuous-flow measurements were performed using a homemade instrument of design and methodology similar to that published by Shastry and Roder,[30] as described in Refs. 31 and 32.

References

1. Barrick D (2009) What have we learned from the studies of two-state folders, and what are the unanswered questions about two-state protein folding? Phys Biol 6: 015001–015009.
2. Jackson SE,Fersht AR (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. Biochemistry 30:10428–10435.
3. Jackson SE (1998) How do small single-domain proteins fold? Fold Des 3:R81–R91.
4. Nickson AA, Stoll KE, Clarke J (2008) Folding of a LysM domain: entropy-enthalpy compensation in the transition state of an ideal two-state folder. J Mol Biol 380:557–569.
5. Cellmer T, Henry ER, Hofrichter J, Eaton WA (2008) Measuring internal friction of an ultrafast-folding protein. Proc Natl Acad Sci USA 105:18320–18325.
6. Cellmer T, Henry ER, Kubelka J, Hofrichter J, Eaton WA (2007) Relaxation rate for an ultrafast folding protein is independent of chemical denaturant concentration. J Am Chem Soc 129:14564–14565.
7. White GW, Gianni S, Grossmann JG, Jemth P, Fersht AR, Daggett V (2005) Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. J Mol Biol 350:757–775.
8. Parker MJ, Spencer J, Clarke AR (1995) An integrated kinetic analysis of intermediates and transition states in protein folding reactions. J Mol Biol 253:771–786.
9. Gianni S, Ivarsson Y, Jemth P, Brunori M, Travaglini-Allocatelli C (2007) Identification and characterization of protein folding intermediates. Biophys Chem 128:105–113.
10. Khorasanizadeh S, Peters ID, Roder H (1996) Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. Nat Struct Biol 3:193–205.
11. Sauder JM, MacKenzie NE, Roder H (1996) Kinetic mechanism of folding and unfolding of Rhodobacter capsulatus cytochrome c2. Biochemistry 35:16852–16862.
12. Fersht A (1999) Structure and Mechanism in Protein Science. New York: Freeman.
13. Krantz BA, Mayne L, Rumbley J, Englander SW, Sosnick TR (2002) Fast and slow intermediate accumulation and the initial barrier mechanism in protein folding. J Mol Biol 324:359–371.
14. Lee CW, Arai M, Martinez-Yamout MA, Dyson HJ, Wright PE (2009) Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP. Biochemistry 48:2115–2124.
15. Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. Cell 91:741–752.
16. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447:1021–1025.
17. Zor T, De Guzman RN, Dyson HJ, Wright PE (2004) Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. J Mol Biol 337: 521–534.
18. Horng JC, Tracz SM, Lumb KJ, Raleigh DP (2005) Slow folding of a three-helix protein via a compact intermediate. Biochemistry 44:627–634.
19. Tollinger M, Kloiber K, Agoston B, Dorigoni C, Lichtenecker R, Schmid W, Konrat R (2006) An isolated helix persists in a sparsely populated form of KIX under native conditions. Biochemistry 45:8885–8893.
20. Friel CT, Smith DA, Vendruscolo M, Gsponer J, Radford SE (2009) The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. Nat Struct Mol Biol 16:318–324.
21. Ferguson N, Capaldi AP, James R, Kleanthous C, Radford SE (1999) Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. J Mol Biol 286:1597–1608.
22. Jemth P, Johnson CM, Gianni S, Fersht AR (2008) Demonstration by burst-phase analysis of a robust folding intermediate in the FF domain. Protein Eng Des Sel 21:207–214.
23. Qi PX, Sosnick TR, Englander SW (1998) The burst phase in ribonuclease A folding and solvent dependence of the unfolded state. Nat Struct Biol 5:882–884.
24. Silow M, Oliveberg M (1997) Transient aggregates in protein folding are easily mistaken for folding intermediates. Proc Natl Acad Sci USA 94:6084–6086.
25. Otzen DE, Kristensen O, Proctor M, Oliveberg M (1999) Structural changes in the transition state of protein folding: alternative interpretations of curved chevron plots. Biochemistry 38:6499–6511.
26. Capaldi AP, Shastry MC, Kleanthous C, Roder H, Radford SE (2001) Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. Nat Struct Biol 8:68–72.
27. Gianni S, Travaglini-Allocatelli C, Cutruzzolà F, Brunori M, Shastry MC, Roder H (2003) Parallel pathways in cytochrome c(551) folding. J Mol Biol 330: 1145–1152.
28. Jemth P, Gianni S, Day R, Li B, Johnson CM, Daggett V, Fersht AR (2004) Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by

kinetics, protein engineering, and simulation. Proc Natl Acad Sci USA 101:6450–6455.
29. Park SH, Shastry MC, Roder H (1999) Folding dynamics of the B1 domain of protein G explored by ultrarapid mixing. Nat Struct Biol 6:943–947.
30. Shastry MC, Luck SD, Roder H (1998) A continuous-flow capillary mixing method to monitor reactions on the microsecond time scale. Biophys J 74:2714–2721.
31. Gianni S, Ivarsson Y, Bah A, Bush-Pelc LA, Di Cera E (2007) Mechanism of Na(+) binding to thrombin resolved by ultra-rapid kinetics. Biophys Chem 131:111–114.
32. Gianni S, Walma T, Arcovito A, Calosci N, Bellelli A, Engström A, Travaglini-Allocatelli C, Brunori M, Jemth P, Vuister GW (2006) Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. Structure 14:1801–1809.

Biochemical and Biophysical Research Communications 428 (2012) 205–209

Contents lists available at SciVerse ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

A folding-after-binding mechanism describes the recognition between the transactivation domain of c-Myb and the KIX domain of the CREB-binding protein

Stefano Gianni *, Angela Morrone, Rajanish Giri, Maurizio Brunori *

Istituto Pasteur-Fondazione Cenci Bolognetti and Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli", Sapienza Università di Roma, Piazzale A. Moro 5, 00185 Rome, Italy

ARTICLE INFO

Article history:
Received 14 September 2012
Available online 28 September 2012

Keywords:
Intrinsically unstructured proteins
Reaction mechanism
Kinetics
Folding
Protein–protein recognition

ABSTRACT

A large body of evidence suggests that a considerable fraction of the human proteome may be at least in part intrinsically unstructured. While disordered, intrinsically unstructured proteins are nevertheless functional and mediate many interactions. Despite their significant role in regulation, however, little is known about the molecular mechanism whereby intrinsically unstructured proteins exert their function. This basic problem is critical to establish the role, if any, of disorder in cellular systems. Here we present kinetic experiments supporting a mechanism of binding-induced-folding when the KIX domain of the CREB-binding protein binds the transactivation domain of c-Myb, an intrinsically unstructured domain. The high-resolution structure of this physiologically important complex was previously determined by NMR spectroscopy. Our data reveal that c-Myb recognizes KIX by first forming a weak encounter complex in a disordered conformation, which is subsequently locked-in by a folding step, i.e. binding precedes folding. On the basis of the pH dependence of the observed combination and dissociation rate constants we propose a plausible mechanism for complex formation. The implications of our results in the light of previous work on intrinsically unstructured systems are discussed.

1. Introduction

Much of our current understanding of proteins is based on the structure-function dogma, 'sequence determines structure determines function'. The fairly recent discovery that a significant fraction of proteins lacks a well-defined structure under native conditions and is intrinsically disordered (known as Intrinsically Denatured Proteins, IDPs) is set to revolutionize our understanding of protein structure–function relationships [1–7]. Despite their disordered structure, IDPs are often functionally competent being engaged in complex mechanisms of protein–protein or protein–DNA recognition. Moreover, IDPs when bound to targets undergo a structural transition to a folded state. The scientific community has been recently challenged to reconcile the unexpectedly high fraction of IDPs with the structure–function dogma. Why are some proteins disordered under native conditions, while others are folded? Is there any potential value for a protein to display a flexible structure that can be so extreme as to compromise the native state?

There are several hypotheses on the significance and value of disorder, which demand experimental validation [5]. From a thermodynamic viewpoint, it has been suggested that destabilization of a native fold may lower the affinity of a protein for its ligand, without necessarily compromising specificity [6,8]. Also, an interesting mechanistic model proposed by Wolynes and co-workers suggests the IDPs to display an increased capture radius to recruit and bind partners [9]. According to this view, called fly-casting mechanism, a disordered protein should form with its physiological partner a high-energy complex that would be locked in place by the coupled folding reaction. Whilst the kinetic data of binding and recognition of IDPs to their physiological partners are relatively scarce, it is of critical importance to understand the mechanisms whereby disordered systems operate [10,11].

The CREB-binding protein (CBP) is a co-activator that mediates the interaction between DNA-bound activator proteins and the components of the basal transcription complex. A well known globular domain of CBP, known as KIX, modulates such interactions [12]. Despite its small size, 87 amino acids, and a relatively simple fold, the KIX domain binds different IDP systems. Two different binding sites mediate such interactions, called c-Myb and MLL sites (named after two characteristic ligands of each of the site, i.e. the transactivation domain of the protein c-Myb and the mixed lineage leukemia MLL protein) [13]. Whilst the structural features of the interaction between KIX and different IDPs have been investigated by NMR spectroscopy in detail and several sets of co-ordinates are available in the pdb database [12–16], the

* Corresponding authors.
E-mail addresses: stefano.gianni@uniroma1.it (S. Gianni), maurizio.brunori@uniroma1.it (M. Brunori).

0006-291X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2012.09.112

mechanism of recognition is still a conundrum. More specifically, it is unclear whether KIX recognizes the IDPs in their ordered or disordered conformation – a critical question in tackling the proposition of an added value for disorder in IDP systems [10,17].

In this paper we present an extensive kinetic study to unveil the mechanism of binding and recognition between KIX and the transactivation domain of c-Myb, an IDP protein that folds into a helical structure once bound to KIX [15]. Our results clearly demonstrate that c-Myb recognizes KIX in an unstructured conformation and that binding precedes folding. Furthermore, by studying the binding kinetics as a function of pH, we propose a plausible recognition mechanism for complex formation.

2. Materials and methods

2.1. Site-directed mutagenesis and protein expression and purification

The site-directed mutant Y72W named pwtKIX was obtained by using the QuikChange mutagenesis kit (Stratagene) according to the manufacturer's instructions. The mutation was confirmed by DNA sequencing. Proteins were purified by using a nickel(II)-charged chelating Sepharose FF (Amersham Biosciences) column equilibrated with 40 mM Tris–HCl and 500 mM NaCl, pH 8.5. The His-tagged wtKIX and pwtKIX were eluted with 250 mM imidazole. The samples were then diluted 4-fold in 40 mM Tris, pH 8.5 and all minor impurities were removed by the purification step on a Q-Sepharose column equilibrated with 40 mM Tris, pH 8.5. The proteins passed through the Q-column; the flow-through containing the protein was collected and concentrated. The purity of the protein was confirmed by SDS–PAGE.

The engineered construct indicated as c-Myb* where c-Myb was fused to the pro-domain of subtilisin was cloned into pPAL7 vector (BioRad). The protein was purified by using a cation-exchange chromatography (S-Sepharose column equilibrated with 40 mM Tris–HCl pH 8.5). c-Myb* was eluted with 850 mM NaCl. The sample was then diluted 4-fold in 40 mM Tris, pH 8.5 and all minor impurities were removed by the purification step on a nickel(II)-charged chelating Sepharose FF (Amersham Biosciences) column equilibrated with 40 mM Tris–HCl equilibrated pH 8.5. The protein passed through the nickel(II)-column; the flow-through containing the protein was collected and concentrated. The purity of the protein was confirmed by SDS–PAGE.

2.2. Equilibrium experiments

2.2.1. Circular dichroism (CD)

Far-UV CD spectra were recorded between 250 and 200 nm at a concentration of 10 μM using a Jasco spectropolarimeter (Jasco, Inc., Easton, MD, USA) and a 1-cm pathlength quartz cuvette (Hellma, Plainview, NY, USA). Data were fitted to a standard 2-state transition. The buffer used was 50 mM sodium phosphate pH 7.2.

2.2.2. Fluorescence

Equilibrium binding experiments were carried on a Fluoromax single photon counting spectrofluorometer (Jobin–Yvon, NJ, USA). Tryptophan fluorescence emission spectra were recorded in a cuvette (1 cm light path) between 300 and 400 nm. The excitation wavelength was 280 nm. pwtKIX concentration was typically 1 μM.

2.3. Stopped-flow measurements

Single mixing kinetic binding experiments were carried out on a SX-18 stopped-flow instruments (Applied Photophysics, Leatherhead, UK); the excitation wavelength was 280 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter. The experiments were performed at 10 °C and the buffer used were: 50 mM Tris/HCl from pH 9.0 to 8.0, 50 mM sodium phosphate from pH 8.0 to 6.3, 50 mM sodium acetate from pH 5.5 to 3.8. All reagents were of analytical grade.

2.4. Temperature-jump fluorescence spectroscopy

The relaxation kinetics was measured as a function of [c-Myb] in absence or in presence of trifluoroethanol (TFE) by using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (Hi-Tech, Salisbury, UK). Temperature was rapidly changed from 16 to 25 °C with a jump-size of 9 °C. 10–20 individual traces were averaged at any given c-Myb concentrations. The fluorescence change of *N*-acetyltryptophanamide (NATA) was used in control measurements. Degassed and filtered samples were slowly pumped through the 0.5 × 2 mm quartz flow cell before data acquisition. The excitation wavelength was 296 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter.

3. Results

To address the secondary structure of the transactivation domain of c-Myb in solution, we carried out equilibrium circular dichroism (CD) experiments at pH 7.2 and 25 °C. The far-UV CD spectrum of c-Myb, reported in Fig. 1, shows absence of secondary structure, confirming that c-Myb is an IDP [15].

The molecule 2,2,2-Trifluoroethanol (TFE), a well-known helix stabilizer, is known to induce folding in peptides and proteins [18,19]. Because c-Myb is helical once bound to its physiological partner KIX [15], we resorted to use TFE to induce folding in the free state. The CD spectra of c-Myb recorded in the presence of different concentrations of TFE (Fig. 1) show that c-Myb undergoes a folding transition with an apparent midpoint at 18% v/v TFE. The dependence of the CD signal at 222 nm upon [TFE] conforms to a standard two-state transition and allows calculating a folding free energy in water, $\Delta G^0 = 2.5 \pm 0.5$ kcal mol^{-1}.

3.1. Mechanism of binding of c-Myb to KIX

To study the mechanism of recognition between KIX and c-Myb, we produced different constructs characterized by a detectable fluorescence change upon binding. We first introduced a tryptophan residue in the proximity of the binding pocket and produced a pseudo-wild-type Y72W variant, called pwtKIX. The equilibrium binding transition of c-Myb to pwtKIX, monitored by fluorescence, is consistent with a simple hyperbolic behavior, with an apparent K_D of 20 ± 4 μM measured at pH 7.2 in the presence of 150 mM KCl and 25 °C (Fig. 1C).

We studied the kinetics of binding of c-Myb to pwtKIX using a temperature jump discharge capacitor apparatus (TGK Instruments, UK). A mixture containing a constant concentration of pwtKIX (10 μM), incubated with different concentrations of c-Myb, was subjected to a 9 °C rapid temperature jump, with a shift from 16 to 25 °C. Under all conditions, the observed relaxation kinetics were consistent with a single exponential time-course, suggesting the absence of detectable intermediates. The observed dependence of the relaxation rate constant on c-Myb concentration is reported in Fig. 2. Despite the inherent complexity of the folding-and-binding recognition reaction between pwtKIX and c-Myb, the observed relaxation behavior displays a simple second-order linear dependence. Furthermore, the apparent K_D of 30 ± 3 μM that can be calculated from the combination and dissociation rate constants is not inconsistent with the value obtained by equilibrium (K_D of 20 ± 4 μM), supporting the lack of detectable intermediates.

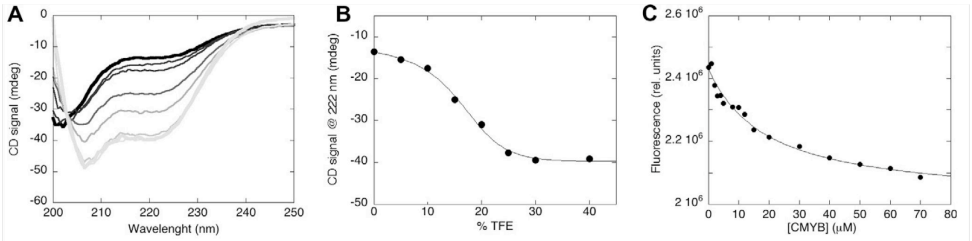


Fig. 1. (A) Far-UV CD spectra of c-Myb measured in 50 mM sodium phosphate buffer containing 150 mM KCl using a 1-cm cylindrical cuvette at pH 7.2 and 10 °C. [c-Myb] = 10 μM in the presence of TFE at increasing concentrations. From top to bottom spectra are depicted in a gray scale varying from black (in the absence of TFE) to light gray (in the presence of 40% TFE). (B) Dependence on [TFE] of the CD signal of c-Myb monitored at 222 nm in 50 mM sodium phosphate buffer, 150 mM KCl at pH 7.2 and 10 °C. Line is the best fit to a standard two-state transition. (C) Equilibrium binding transition of c-Myb to pwtKIX by fluorescence quenching, as observed in 50 mM sodium phosphate buffer, 150 mM KCl at pH 7.2. The emission wavelength was 370 nm (excitation 280 nm). pwtKIX concentration was 1 μM. Line is the best fit to a simple hyperbolic equation.

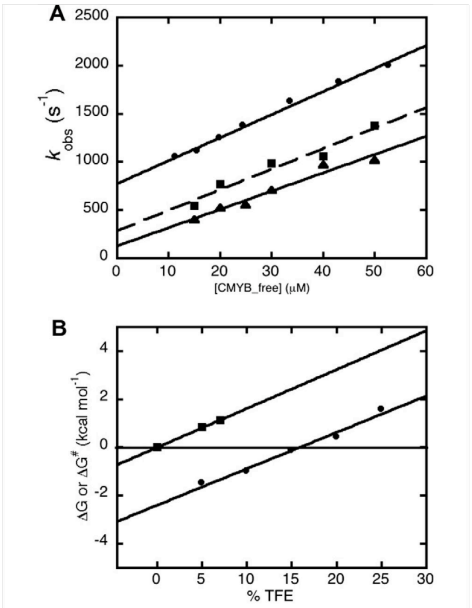
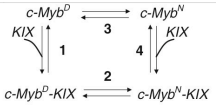


Fig. 2. (A) Dependence on the concentration of c-Myb of the relaxation rate constant measured by T-jump in 50 mM sodium phosphate buffer,150 mM KCl at pH 7.2, in the absence (●) and in the presence of 5% TFE (■) and 7% TFE (▲). Temperature was rapidly changed from 16 °C to 25 °C. (B) Dependence on [TFE] of the activation energy for dissociation of the c-Myb–pwtKIX complex (■), compared to the effect on the equilibrium free energy of folding of c-Myb (●), as calculated from the titration reported in Fig. 1.

To shed light on the function, if any, of disorder in IDP systems would require to clarify the order of events between folding and binding: does folding occur prior binding or do IDPs recognize their partner in a disordered conformation and fold only after binding? A monomeric protein undergoing a ligand induced conformational change can be described by a square mechanism (Scheme 1). Here c-Myb^D and c-Myb^N denote the denatured and folded



Scheme 1.

conformations of c-Myb, respectively. Complex formation progressing through pathway 1 and 2 is representative of an induced-fit model, whereby ligand binding induces a conformational change [20], i.e. folding-after-binding. Alternatively, binding progressing through pathway 3 and 4 assumes that two alternative conformations of c-Myb are in pre-equilibrium in the absence of the ligand, formally similar to a concerted model [21], or folding-before-binding.

To distinguish between the two different scenarios, we performed T-jump experiments in the presence of increasing concentrations of TFE, which stabilizes the helical conformation of c-Myb (Fig. 1). Remarkably, inspection of Fig. 2A reveals that, while the presence of TFE has a negligible effect on the apparent association rate constant, it stabilizes the c-Myb–KIX complex by lowering the dissociation rate constant. In fact, although experience suggests estimates of dissociation rate constants from pseudo-first-order dependences could be prone to large experimental errors, it is evident that, while TFE slows down binding, the slope of the second order plot is nearly unchanged. In analogy to the so called Φ-value analysis in protein folding studies [22], the lack of stabilization of the rate limiting step suggests the folding transition state to be as unstructured as the denatured c-Myb. Consequently, it may be concluded that folding of c-Myb is a late event occurring rapidly after the main rate-limiting barrier and being associated to the observed fluorescence quenching (Fig. 1), the overall reaction being consistent with a folding-after-binding mechanism. In these experiments, the highest concentration of TFE employed was 7% (v/v) because we found pwtKIX to precipitate at higher concentrations.

A powerful method to infer quantitatively the effect of a perturbation on the reaction mechanism is to compare the effect on activation free energy with that on equilibrium [23]. Fig. 2B depicts the dependence on the [TFE] of the activation energy for the dissociation of the c-Myb–KIX complex, compared to the effect on the equilibrium free energy of folding of c-Myb, as calculated from the TFE titration reported in Fig. 1. The linear dependence of the two parameters on [TFE] is essentially the same, while the association rate constant is essentially insensitive to [TFE] (Fig. 2A). This finding strongly suggests that the stabilization of c-Myb in its

helical folded conformation may account for all the stability the c-Myb–KIX complex. Furthermore, because the stabilization is due to a decrease of the dissociation rate constant, it may be concluded that the folding of c-Myb is a late step that occurs only downhill the main limiting barrier, i.e. c-Myb recognizes KIX in a disordered non-helical conformation.

3.2. Effect of pH on the binding kinetics

In order to produce c-Myb in high yields using *Escherichia coli* expression strains, we engineered a construct where the transactivation domain of c-Myb was fused to the pro-domain of subtilisin, hereby denoted as c-Myb*. While this construct binds to pwtKIX with an overall affinity comparable to that of free c-Myb, an investigation of its binding kinetics (Fig. S1) revealed a clear-cut effect on both the association and dissociation rate constants. These relatively limited changes are most likely due to some steric hindrance induced by the presence of the large tag, but do not affect the K_D for the binding to KIX. Furthermore, c-Myb* displayed the additional advantage of a much better signal to noise ratio compared to free c-Myb. Because of such improved signal to noise due to the tag, we resorted to expand the analysis of the mechanism of binding using c-Myb* and explored the effect of pH from 4 to 9.

The dependence of the observed rate constant on the concentration of c-Myb*, depicted in Fig. S2, shows that pH affects both the intercept and the slope, the overall behavior nevertheless being consistent with a simple bimolecular reaction. To validate the pH dependence of the apparent dissociation rate constant, we carried out a full set of displacement kinetic experiments where a pre-incubated complex between pwtKIX and c-Myb* was rapidly mixed with an excess of wild-type KIX (i.e. the wild type domain with a Tyr at position 72). Under appropriate conditions, the reaction is rate limited by the dissociation of the pre-existing complex, the excess of wild-type KIX substituting the bound pwtKIX in the complex with c-Myb*. The dissociation process was measured at different concentrations of wild-type KIX, ranging from 2- to 10-fold excess; the observed rate constants were found to be insensitive to wild-type KIX concentration at all pH values (Fig. S3). This methodology allows estimating unequivocally the dissociation rate constant, in analogy to classical experiments on myoglobin [24].

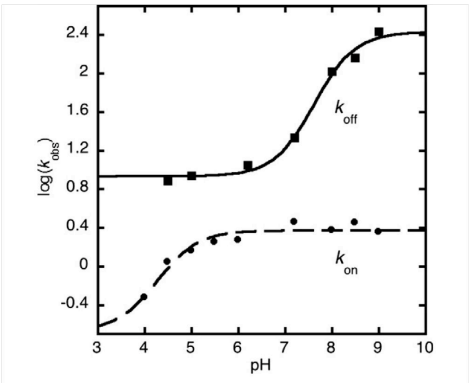


Fig. 3. Dependence of the logarithm of the apparent association (●) and dissociation (■) rate constants on pH. The apparent dissociation rate constants were obtained by a displacement experiment involving a rapid mixing of a pre-incubated complex between pwtKIX-c-Myb* with an excess of wtKIX, measured at 10 °C. Lines are the best fit to an equation implying a single protonation site.

A plot of the dependence of the apparent association and dissociation rate constants on pH is reported in Fig. 3. Both rate constants increase with pH following a sigmoidal profile, consistent with the protonation of a single group; the apparent pKa's for the association and dissociation rate constants are 4.2 and 7.6, respectively.

4. Discussion

The mechanism of recognition between IDPs and their physiological partners is expected to be a complex reaction that involves (i) the productive encounter between the two partners, (ii) the folding of the IDP system and (iii) the locking of key stabilizing interactions [1–7]. Despite the frequent occurrence of disorder in the proteome, its role is still shrouded in mystery, and even the order of these three key events in a folding-and-binding reaction is still unclear [10]. Hagen and co-workers [25] recently provided a complete characterization of the binding kinetics of the IDP protein IA3 to the yeast aspartic proteinase A; it was found that IA3 binds in a disordered conformation, a recognition event preceding folding. On the other hand, Onitsuka et al. [26] investigated the folding and binding reaction of disordered mutants of staphylococcal nuclease, which fold upon ligand binding. Interestingly, different variants seemed to follow either the folding-after-binding or the folding-before-binding mechanism, suggesting that the energetic partitioning between the different paths shown in Scheme 1 may be fine tuned via mutagenesis, as implicit in a squared model.

In a seminal work, Wright and co-workers [27] described the structural features of the interaction between the KIX domain and the IDP system pKID. Remarkably, by equilibrium NMR, it was found that partially folded species of pKID may bind to KIX, suggesting that, also in this case, the most plausible mechanism of IDP recognition implies a binding event preceding folding. Their equilibrium data, however, while providing the finest structural details of the interaction, could not unequivocally establish whether the recognition mechanism was a folding-after-binding rather than a folding-before-binding, a task that demands kinetic experiments. Identification of an intermediate by equilibrium measurements is not diagnostic on its role in the reaction mechanism, since it may be either on- or off- the productive pathway [28].

We report hereby an extensive kinetic characterization of the recognition reaction between KIX and c-Myb. Despite the inherent complexity of the process, it appears that, under all conditions explored, folding and binding are coupled and highly co-operative. Indeed no transient intermediates could be detected by fluorescence spectroscopy, the time course of the T-jump induced relaxation being described by a single exponential; and likewise the dissociation time course obtained by displacement was, under all conditions, consistent with a single exponential process. Comparison of the data recorded in the presence and absence of TFE, which stabilizes the helical conformation of c-Myb, sheds light on the order of events in the ligand induced folding. In fact, because TFE affects only the dissociation rate constant without influence on the association rate constant, it may be concluded that in the rate limiting step c-Myb is essentially unstructured. Therefore, binding precedes folding.

It is of interest to analyze the pH dependence of the observed association and dissociation rate constants (Fig. 3), whose pH dependence is consistent with the titration a single group with pKa of 4.2 and 7.6 respectively. Examination of the structure of the complex between folded c-Myb and KIX [15] (Fig. 4) reveals the presence at the interface of a single His residue on KIX (His17) forming a hydrogen bond with Thr305 of c-Myb, clearly compatible with the observed pKa of 7.6. Furthermore, two acidic residues at the interface, namely Glu306 and Glu308 of c-Myb

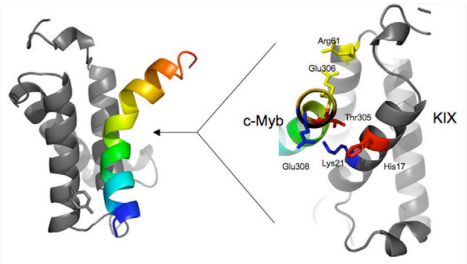


Fig. 4. Three dimensional structure of the complex between KIX and c-Myb, from the NMR data published by Wright and co-workers [15]. Residue Tyr72, mutated to Trp in pwtKIX, is highlighted in sticks on the structure on the left. The interaction surface, as seen from the right (see the arrow), is zoomed on the right of the figure. Amino acid side-chains highlighted are: His17-Thr305 (red); Lys21-Glu306 (blue); Arg61-Glu308 (yellow). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

form salt bridges with Arg61 and Lys21 of KIX, respectively. Both residues may be compatible with the pKa value of 4.2 controlling the association rate constant and could, therefore, be responsible for its pH dependence. We note, however, that while the Lys21-Glu308 salt bridge is adjacent to the His17-Thr305 interaction, the Arg61-Glu306 is located at the opposite side of the binding interface. Therefore, because the effects on the association and dissociation rate constants are distinct, we suggest Glu306 to be responsible for the pH dependence of the association rate constant. Overall, our data are consistent with a possible scenario whereby the rate limiting transition state is characterized by a largely unstructured conformation of c-Myb, which binds initially to KIX via a salt bridge between Arg61 and Glu306. The subsequent rapid folding step is driven by the locking interactions at the opposite side of the binding interface, involving the hydrogen bond between His17 and Thr305. In analogy with these observations, NMR relaxation dispersion on the KIX–pKID complex suggested the electrostatic interactions between His17 of KIX and Asp140/Asp144 of pKID to be formed at the latest stages of binding, being absent in the partially folded intermediate [27]. This finding would suggest the mechanism of recognition between KIX and different IDPs to display some conserved features. Future work based on Φ-value analysis will shed light on the atomistic details of the folding and binding reactions involved in the recognition and complex formation between KIX and its different IDP partners.

Acknowledgments

Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (RBRN07BMCT_007), to M.B. and (PNR-CNR Aging Program 2012-2014) to S.G.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2012.09.112.

References

[1] A.K. Dunker, I. Silman, V.N. Uversky, J.L. Sussman, Function and structure of inherently disordered proteins, Curr. Opin. Struct. Biol. 18 (2008) 756–764.
[2] H.J. Dyson, P.E. Wright, Coupling of folding and binding for unstructured proteins, Curr. Opin. Struct. Biol. 12 (2002) 54–60.
[3] P. Tompa, The interplay between structure and function in intrinsically unstructured proteins, FEBS Lett. 579 (2005) 3346–3354.
[4] P. Tompa, Unstructural biology coming of age, Curr. Opin. Struct. Biol. 21 (2011) 419–425.
[5] V.N. Uversky, A.K. Dunker, Understanding protein non-folding, Biochim. Biophys. Acta 1804 (2010) 1231–1264.
[6] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm, J. Mol. Biol. 293 (1999).
[7] P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions, Trends Biochem. Sci. 33 (2008) 2–8.
[8] R.S. Spolar, M.T.J. Record, Coupling of local folding to site-specific binding of proteins to DNA, Science 263 (1994) 777–784.
[9] B.A. Shoemaker, J.J. Portman, P.G. Wolynes, Speeding molecular recognition by using the folding funnel: the fly-casting mechanism, Proc. Nat. Acad. Sci. USA 97 (2000) 8868–8873.
[10] T. Kiefhaber, A. Bachmann, K.S. Jensen, Dynamics and mechanisms of coupled protein folding and binding reactions, Curr. Opin. Struct. Biol. 22 (2012) 21–29.
[11] J. Dogan, T. Schmidt, X. Mu, A. Engstrom, P. Jemth, Fast association and slow transitions in the interaction between two intrinsically disordered protein domains, J. Biol. Chem. (2012) (Epub ahead of print).
[12] I. Radhakrishnan, G.C. Pérez-Alvarado, D. Parker, H.J. Dyson, M.R. Montminy, P.E. Wright, Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions, Cell 91 (1997) 741–752.
[13] N.K. Goto, T. Zor, M. Martinez-Yamout, H.J. Dyson, P.E. Wright, Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain, J. Biol. Chem. 277 (2002) 43168–43174.
[14] C.W. Lee, M. Arai, M.A. Martinez-Yamout, H.J. Dyson, P.E. Wright, Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP, Biochemistry 48 (2009) 2115–2124.
[15] T. Zor, R.N. De Guzman, H.J. Dyson, P.E. Wright, Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb, J. Mol. Biol. 337 (2004) 521–534.
[16] F. Wang, C.B. Marshall, K. Yamamoto, G.Y. Li, G.M. Gasmi-Seabrook, H. Okada, T.W. Mak, M. Ikura, Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment, Proc. Nat. Acad. Sci. USA 109 (2010) 6078–6083.
[17] T.R. Weikl, C. von Deuster, Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis, Proteins 75 (2009) 104–110.
[18] A. Jasanoff, A.R. Fersht, Quantitative determination of helical propensities from trifluoroethanol titration curves, Biochemistry 33 (1994) 2126–2135.
[19] J.K. Myers, C.N. Pace, J.M. Scholtz, Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1, Protein Sci. 7 (1998) 383–388.
[20] D.E.J. Koshland, G. Némethy, D. Filmer, Comparison of experimental binding data and theoretical models in proteins containing subunits, Biochemistry 5 (1966) 365–385.
[21] J. Monod, J. Wyman, J.P. Changeux, On the nature of allosteric transitions: a plausible model, J. Mol. Biol. 12 (1965) 88–118.
[22] A.R. Fersht, A. Matouschek, L. Serrano, The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding, J. Mol. Biol. 224 (1992) 771–782.
[23] J.E. Leffler, Parameters for the description of transition states, Science 117 (1953) 340–341.
[24] E. Antonini, M. Brunori, Hemoglobin and Myoglobin in Their Reactions with Ligands, North-Holland Publ Co., Amsterdam, The Netherlands, 1971.
[25] R. Narayanan, O.K. Ganesh, A.S. Edison, S.J. Hagen, Kinetics of folding and binding of an intrinsically disordered protein: the inhibitor of yeast aspartic proteinase YPrA, J. Am. Chem. Soc. 130 (2008) 11477–11485.
[26] M. Onitsuka, H. Kamikubo, Y. Yamazaki, M. Kataoka, Mechanism of induced folding: both folding before binding and binding before folding can be realized in staphylococcal nuclease mutants, Proteins 72 (2008) 837–847.
[27] K. Sugase, H.J. Dyson, P.E. Wright, Mechanism of coupled folding and binding of an intrinsically disordered protein, Nature 447 (2007) 1021–1025.
[28] A. Fersht, Structure and Mechanism in Protein Science, W.H. Freeman and Co., New York, 1999.